(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 October 2004 (28.10.2004)

PCT

(10) International Publication Number
WO 2004/092360 A2

(51) International Patent Classification[7]: C12N 7/00

(21) International Application Number: PCT/US2004/011710

(22) International Filing Date: 9 April 2004 (09.04.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/462,218	10 April 2003 (10.04.2003)	US
60/462,465	11 April 2003 (11.04.2003)	US
60/462,418	12 April 2003 (12.04.2003)	US
60/462,748	13 April 2003 (13.04.2003)	US
60/463,109	14 April 2003 (14.04.2003)	US
60/463,460	15 April 2003 (15.04.2003)	US
60/463,668	16 April 2003 (16.04.2003)	US
60/463,983	17 April 2003 (17.04.2003)	US
60/463,971	18 April 2003 (18.04.2003)	US
60/464,899	22 April 2003 (22.04.2003)	US
60/464,838	22 April 2003 (22.04.2003)	US
60/465,273	23 April 2003 (23.04.2003)	US
60/465,535	24 April 2003 (24.04.2003)	US
60/468,312	5 May 2003 (05.05.2003)	US
60/473,144	22 May 2003 (22.05.2003)	US
60/495,024	14 August 2003 (14.08.2003)	US
60/505,652	23 September 2003 (23.09.2003)	US
60/510,781	11 October 2003 (11.10.2003)	US
60/529,464	11 December 2003 (11.12.2003)	US
60/536,177	12 January 2004 (12.01.2004)	US
60/560,757	7 April 2004 (07.04.2004)	US

(71) Applicant *(for all designated States except US)*: **CHIRON CORPORATION** [US/US]; 4560 Horton Street, Emeryville, CA 94608-2916 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **RAPPUOLI, Rino** [IT/IT]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **MASIGNANI, Vega** [IT/IT]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **STADLER, Konrad** [DE/DE]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **GREGERSEN, Jens-Peter** [DE/DE]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **CHIEN, David** [US/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **HAN, Jang** [US/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **POLO, John** [US/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **WEINER, Amy** [US/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **HOUGHTON, Michael** [GB/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **SONG, Hyun, Chul** [KR/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **SEO, Mi, Young** [KR/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **DONNELLY, John, J.** [US/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **KLENK, Hans, Dieter** [DE/DE]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **VALIANTE, Nicholas** [US/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(74) Agents: **HALE, Rebecca, M.** et al.; Chiron Corporation, Intellectual Property R338, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: THE SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS

(57) Abstract: An outbreak of a virulent respiratory virus, now known as Severe Acute Respiratory Syndrome (SARS), was identified in Hong Kong, China and a growing number of countries around the world in 2003. The invention relates to nucleic acids and proteins from the SARS coronavirus. These nucleic acids and proteins can be used in the preparation and manufacture of vaccine formulations, diagnostic reagents, kits, etc. The invention also provides methods for treating SARS by administering small molecule antiviral compounds, as well as methods of identifying potent small molecules for the treatment of SARS.

WO 2004/092360 A2

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

THE SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS

All documents cited herein are incorporated by reference in their entirety.

RELATED APPLICATIONS, FROM WHICH PRIORITY IS CLAIMED

This application incorporates by reference in its entirety US provisional patent application 60/462,218, Attorney Reference No. PP20474.001, filed on April 10, 2003 via Express Mail with the US post office, US provisional patent application 60/462,465, Attorney Reference No. PP20480.001, filed on April 11, 2003 via Express Mail with the US post office, US provisional patent application 60/462,418, Attorney Reference No. PP20480.002, filed on April 12, 2003 via Express Mail with the US post office, US provisional patent application 60/462,748, Attorney Reference No. PP20480.003, filed on April 13, 2003 via Express Mail with the US post office, US provisional patent application 60/463,109, Attorney Reference No. PP20480.004, filed on April 14, 2003 via Express Mail with the US post office, US provisional patent application 60/463,460, Attorney Reference No. PP20480.005, filed on April 15, 2003 via Express Mail with the US post office, US provisional patent application 60/463,668, Attorney Reference No. PP20480.006, filed on April 16, 2003 via Express Mail with the US post office, US provisional patent application 60/463,983, Attorney Reference No. PP20480.007, filed on April 17, 2003 via Express Mail with the US post office, US provisional patent application 60/463,971, Attorney Reference No. PP20480.008, filed on April 18, 2003 via Express Mail with the US post office, US provisional patent application 60/464,899, Attorney Reference No. PP20480.009, filed on April 22, 2003 via Express Mail with the US post office, US provisional patent application 60/464,838, Attorney Reference No. PP20507.001, filed on April 22, 2003 via Express Mail with the US post office, US provisional patent application 60/465,273, Attorney Reference No. PP20518.001, filed on April 23, 2003 via Express Mail with the US post office, US provisional patent application 60/465,535, Attorney Reference No. PP20518.002, filed on April 24, 2003 via Express Mail with the US post office, US provisional patent application 60/468,312, Attorney Reference No. PP20480.010, filed on May 5, 2003 via Express Mail with the US post office, and US provisional patent application 60/473,144, Attorney Reference No. PP20480.011, filed on May 22, 2003, US provisional patent application 60/495,024, Attorney Reference No. PP20480.012, filed on August 14, 2003 via Express Mail with the US post office, US provisional patent application 60/505,652, Attorney Reference No. PP20480.013, filed on September 23, 2003 via Express Mail with the US post office, US provisional patent application 60/510,781, Attorney Reference No. PP20480.014, filed on October 11, 2003 via Express Mail with the US

post office, US provisional patent application 60/529,464, Attorney Reference No. PP20480.015, filed on December 11, 2003 via Express Mail with the US post office, US provisional patent application 60/536,177, Attorney Reference No. PP20480.016, filed on January 12, 2004 via Express Mail with the US post office, and US provisional patent application 60/___,___, Attorney Reference No. PP20480.017, filed on April 7, 2004 via Express Mail with the US post office.

FIELD OF THE INVENTION

The invention relates to nucleic acids and proteins from Severe Acute Respiratory Syndrome (SARS) Virus. These nucleic acids and proteins can be used in the preparation and manufacture of vaccine formulations for the treatment or prevention of SARS. The invention also relates to diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention also relates to methods for the treatment or prevention of SARS utilizing small molecule viral inhibitors and combinations of small molecule viral inhibitors and kits for the treament of SARS.

BACKGROUND OF THE INVENTION

An outbreak of a virulent respiratory virus, now known as Severe Acute Respiratory Syndrome (SARS), was identified in Hong Kong, China and a number of other countries around the world in 2003. Patients typically had symptoms including fever, dry cough, dyspnea, headache, and hypoxemia. Isolates of the SARS virus appear to have homology with at least the RNA polymerase gene of several known coronaviruses. A phylogenetic analysis of this homology is presented in Peiris *et al.*, "Coronavirus as a possible cause of severe acute respiratory syndrome", *Lancet*, published online April 8, 2003 at *http://image.thelancet.com/extras/03art3477web.pdf*, incorporated herein by reference in its entirety. Other sequenced fragments of the SARS virus genome appear to overlap with the open reading frame 1b of coronaviruses. See, Drosten *et al.*, "Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome", *New England Journal of Medicine*, published online at *http://www.nejm.org* on April 10, 2003, incorporated herein by reference in its entirety.

The Genome Science Center in British Colombia, Canada published on its website (*http://www.bcgsc.ca/bioinfo/SARS/*) a draft genome assembly of 29,736 base pairs of a virus believed to be a SARS virus, referred to as the TOR2 isolate. This draft genome assembly is given herein as SEQ ID NO: 1.

The Centers for Disease Control (CDC) published a nucleotide sequence of a SARS-CoV strain (SEQ ID NO: 2) on its website (*http://www.cdc.gov/ncidod/sars/pdf/nucleoseq.pdf*). The CDC

has also published a phylogenetic tree of the predicted N, S and M proteins (attached as FIGURE 6). This tree places the SARS virus outside any of the previously known coronavirus groups.

There is a growing need for prophylactic or therapeutic vaccines against the SARS virus as well as diagnostic and screening methods and compositions to identify the presence of the virus in, *e.g.*, mammalian tissue or serum.

SUMMARY OF THE INVENTION

The invention relates to nucleic acids and proteins from Severe Acute Respiratory Syndrome (SARS) virus. These nucleic acids and proteins can be used in the preparation and manufacture of vaccine formulations for the treatment or prevention of SARS. Such vaccine formulations may include an inactivated (or killed) SARS virus, an attenuated SARS virus, a split SARS virus preparation and a recombinant or purified subunit formulation of one or more SARS viral antigens. Expression and delivery of the polynucleotides of the invention may be facilitated via viral vectors and/or viral particles.

The invention also relates to diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention further includes non-coding SARS viral polynucleotide sequences, SARS viral sequences encoding for non-immunogenic proteins, conserved and variant SARS viral polynucleotide sequences for use in such diagnostic compositions and methods.

The invention further relates to vaccine formulations comprising one or more SARS virus antigens and one or more other respiratory virus antigens. Additional respiratory virus antigens suitable for use in the invention include antigens from influenza virus, human rhinovirus (HRV), parainfluenza virus (PIV), respiratory syncytial virus (RSV), adenovirus, metapneumovirus, and rhinovirus. The additional respiratory virus antigen could also be from a coronavirus other than the SARS coronavirus. Preferably, the additional respiratory virus antigen is an influenza viral antigen.

The compositions of the invention may further comprise one or more adjuvants. Adjuvants suitable for use in the invention include mucosal, transdermal or parenteral adjuvants. Mucosal adjuvants suitable for use in the invention include detoxified bacterial ADP-ribosylating toxins, such as *E. coli* heat labile toxoids (*e.g.*, LTK63), chitosan and derivatives thereof, and non-toxic double mutant forms of *Bordetella pertussis* toxoids. Parenteral adjuvants suitable for use in the invention include MF59 and aluminum or aluminum salts.

The invention also provides methods for treating SARS by administering small molecule compounds, as well as methods of identifying potent small molecules for the treatment of SARS.

In one aspect of the invention a method of identifying a therapeutically active agent is provided comprising: (a) contacting the therapeutically active agent with a cell infected with the SARS virus; (b) measuring attenuation of a SARS related enzyme.

In a more particular embodiment, the therapeutically active agent is a small molecule. In another more particular embodiment, the therapeutically active agent is a nucleoside analog. In another more particular embodiment the therapeutically active agent is a peptoid, oligopeptide, or polypeptide. In another embodiment the SARS related enzyme is SARS protease. In another embodiment the SARS related enzyme is SARS polymerase. In still another embodiment the SARS related enzyme is a kinase. Methods of identifying therapeutically active agents for treatment of SARS virus infection are further discussed in Section V below.

In another aspect of the invention a method of treating a human infected with SARS is provided comprising administering a small molecule to a patient in need thereof. In one embodiment the small molecule is an inhibitor of SARS protease. In another embodiment the small molecule is an inhibitor of SARS polymerase. In another embodiment the SARS related enzyme is a kinase. In still another embodiment the small molecule is administered orally or parenterally.

The invention also provides the use of such small molecules in the manufacture of a medicament for the treatment of severe acute respiratory syndrome.

Small molecule compounds of the present invention include those of less than 1000 g/mol, preferably with an aromatic region and greater than one heteroatom selected from O, S, or N. Preferred small molecules include, but are not limited to acyclovir, gancyclovir, vidarabidine, foscamet, cidofovir, amantidine, ribavirin, trifluorothymidine, zidovudine, didanosine, zalcitabine, and combinations thereof. Interferons may also be used for treating patients, including interferon-α and interferon-β. Interferon treatment has shown promise in treating SARS in monkeys (Enserink (2004) *Science* 303:1273-1275), particularly when pegylated (Haagmans *et al.* (2004) *Nature Medicine* 10:290-293).

One aspect of the present invention relates to methods for identifying individuals exposed to, and biological samples containing SARS virus (SARSV), and to kits for carrying out the methods. Such methods can utilize nucleic acid detection techniques such as PCR, RT-PCR (the *Coronaviridae* are RNA viruses), transcription-mediated amplification (TMA), ligase chain reaction (LCR), branched DNA signal amplification assays, isothermal nucleic acid sequence based amplification (NASBA), other self-sustained sequence replication assays, boomerang DNA amplification, strand-displacement activation, cycling probe technology, or combinations of such amplification methods. Such nucleic acid detection techniques utilize oligonucleotides having nucleotide sequence similar to, or complementary to, the SARS viral genome, as primers (*e.g.*, for amplification) and as probes (*e.g.*, for capture or detection), as is well known in the art.

Alternatively, or in addition to the nucleic acid detection methods described supra, the methods of the present invention can utilize various immunoassay techniques for detection of SARSV antigens and/or antibodies.

Accordingly, the present invention relates to methods of identifying individuals exposed to SARSV, or biological samples containing SARSV, by detecting the presence of SARSV antigens using antibodies which specifically bind to the same. The antibodies are preferably monoclonal antibodies. Quantification of the amount of viral antigens present in a sample of an individual may be used in determining the prognosis of an infected individual. Preferably, the SARSV antigens to be detected are generally one of the structural proteins, particularly those present on the surface of the viral particles and include, for example, the spike glycoprotein (S), also called E2; the envelope (small membrane) protein (E), also called sM; the membrane glycoprotein (M), also called E1; the hemagglutinin-esterase glycoprotein (HE); also called E3; and the nucleocapsid phosphoprotein (N). In preferred embodiments, the antigens to be detected are the S, E and M proteins using antibodies to the same.

The present invention relates to kits for identifying individual SARSV and reagents used in such kits. The kits comprise a first container which contains antibodies which specifically bind to a SARSV antigen and a second container which contains the SARSV antigen. The antibodies are preferably monoclonal antibodies. The kits may be adapted for quantifying the amount of antigen in a sample of an individual. Such information may be used in determining the prognosis of an infected individual.

The present invention relates to methods of identifying individuals exposed to SARS virus, or biological samples containing SARSV, by detecting the presence of antibodies against SARS virus antigen in a sample using SARS antigen. Quantification of the amount of anti-SARS protein from SARS antibodies present in a sample of an individual may be used in determining the prognosis of an infected individual. Any one or more of the viral proteins (structural proteins or nonstructural proteins) may be used as antigen to detect the SARSV antibodies; preferably a SARSV antigen that is conserved amoung SARSV isolates is preferred. In this regard, nonstructural protein (*e.g.*, Pol, Hel, 3CLp, MP, PLP1, PLP2) may be particularly useful.

The present invention relates to kits for identifying individuals exposed to SARS and reagents used therein. The kits comprise a first container which contains antibodies which were produced in response to exposure to an antigen from SARS virus and a second container which contains the SARS antigen(s). The kits may be adapted for quantifying the amount of anti-SARS antibodies present in a sample of an individual. Such information may be used in determining the prognosis of an infected individual.

The present invention relates to methods of identifying individuals exposed to SARS virus, or biological samples containing SARSV, by detecting the presence of nucleic acid from SARS

virus. Quantification of the amount of SARS nucleic acid present in a sample of an individual may be used in determining the prognosis of an infected individual. The methods utilize oligonucleotide probes and/or primers that are similar or complementary in sequence to the SARSV genome or transcription or replication products. Preferred probes and primers are described herein. Also included in the present invention are kits for carrying out the methods of detecting the SARSV nucleic acid.

The invention further includes a method for the treatment and/or prevention of SARS through the administration of a therapeutically effective amount of at least one antiviral compound from among those described in the US Patents and published international patent applications listed in Table 1 and Table 2. In one embodiment of the method, the antiviral compound is a small molecule. In another embodiment, the antiviral compound is a protease inhibitor. In a further embodiment, the antiviral protease inhibitor is a 3C-like protease inhibitor and/or a papain-like protease inhibitor. In another embodiment, the antiviral compound is an inhibitor of an RNA-dependent RNA polymerase. In another embodiment, a first antiviral compound which is a protease inhibitor is administered with a second antiviral compound which is an RNA-dependent RNA polymerase inhibitor. The invention further provides for the administration of a steroidal anti-inflammatory drug in combination with at least one antiviral compound, for example, from the antiviral compounds described in the documents listed in Table 1 and Table 2.

The invention further provides for a method for the treatment and/or prevention of SARS through the administration of a therapeutically effective amount of at least one antiviral compound from among those described in the US Patents and published international patent applications listed in Table 1 and Table 2 by inhalation. In one embodiment of the method, the antiviral compound is a small molecule. In another embodiment, the antiviral compound is a protease inhibitor. In a further embodiment, the antiviral protease inhibitor is a 3C-like protease inhibitor and/or a papain-like protease inhibitor. In another embodiment, the antiviral compound is an inhibitor of an RNA dependent RNA polymerase. In another embodiment, a first antiviral compound which is a protease inhibitor is administered with a second antiviral compound which is an RNA-dependent RNA polymerase inhibitor. The invention further provides for the administration of a steroidal anti-inflammatory drug in combination with at least one antiviral compound, for example, from the antiviral compounds described in the documents listed in Table 1 and Table 2 by inhalation. The steroidal anti-inflammatory drug may be administered by inhalation for a local effect or administered for systemic absorption such as via an oral or intravenous route.

The invention further provides the use of an antiviral compound, as defined above, in the manufacture of a medicament for the treatment of severe acute respiratory syndrome.

The invention further provides for a kit for use by a consumer for the treatment and/or prevention of SARS. Such a kit comprises: (a) a pharmaceutical composition comprising a therapeutically effective amount of at least one antiviral compound from among those described in the US Patents and published international patent applications listed in Table 1 and Table 2 and a pharmaceutically acceptable carrier, vehicle or diluent; (b) a container for holding the pharmaceutical composition; and, optionally; (c) instructions describing a method of using the pharmaceutical compositions for the treatment and or the prevention of SARS. The kit may optionally contain a plurality of antiviral compounds for the treatment of SARS wherein the anti viral compounds are selected from 3C-like protease inhibitors and papain-like protease inhibitors. In a further embodiment, the kit contains an antiviral compound which is an RNA-dependent RNA polymerase inhibitor. When the kit comprises more than one antiviral compound, the antiviral compounds contained in the kit may be optionally combined in the same pharmaceutical composition.

An additional aspect of the invention provides for the use of at least one of the antiviral compounds described in the US Patents and published international patent applications listed in Table 1 and Table 2 for the manufacture of a medicament for the treatment or prevention of SARS.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1: Schematic of coronavirus genome organization.

FIGURE 2: Schematic of coronavirus ORF1a/ORF1b gene products.

FIGURE 3 (A - C): Alignment of coronavirus polynucleotide sequences for selected genes (including nucleocapsid (N), matrix (M), and hemagluttinin-esterase (HE)).

FIGURE 4 (A - F): Alignment of coronavirus polypeptide sequences (including ORF1a/ORF1b, nucleocapsid (NP), hemagluttinin-esterase (HE), envelope (Sm or E), matrix (M), and spike (S).

FIGURE 5: Alignment of spike (S) polypeptide sequences, taken from Figure 4, in the region of the junction of the S1 and the S2 domains, and protease cleavage site for selected coronaviruses.

FIGURE 6: CDC phylogenetic tree of SARS-CoV strain (Clustalx 1.82, neighbor-joining tree). Figure 6A shows coronavirus N protein analysis, Figure 6B shows coronavirus S protein analysis, and Figure 6C shows coronavirus M protein analysis.

FIGURE 7: Conserved and specific sequence of the SARS virus. Figures 7A-7D show multiple sequence alignments (CLUSTAL W 1.82) of the structural proteins of the SARS virus genome (7A: PEP4 Spike protein; 7B: PEP7 small membrane protein; 7C: PEP8 matrix glycoprotein; 7D: PEP13 nucleocapsid protein), which have counterparts in all or some of the other known coronaviruses. Figures 7E-7H show dendrograms reporting the protein distances among the

sequences in alignments 7A-7D. Labels 229E: human coronavirus; MEV: murine hepatitis virus; TGV: transmissible gastroenteritis virus; AIBV: avian infectious bronchitis virus; BOVINE: Bovine coronavirus; PEDV: porcine epidemic diarrhea virus.

FIGURE 8: Alignment of the 5'UTR of several coronaviruses, to show consensus nucleotide sequence at the 5'UTR.

FIGURE 9: Sequences of preferred primers for amplification of the 5'UTR. F and R denote forward and reverse PCR primers, and the numbers indicate nucleotide positions withing Figure 8.

FIGURE 10: Alignment of the 3'UTR of several coronaviruses, to show consensus nucleotide sequence at the 3'UTR.

FIGURE 11: Sequences of preferred primers for amplification of the 3'UTR. F and R denote forward and reverse PCR primers, and numbers indicate nucleotide positions within Figure 10.

FIGURE 12: Coiled-coil prediction for SEQ ID NO: 6042, using Coils program (Figure 12A) or LearrCoil (Figure 12B).

FIGURE 13: Example of insertion of a reporter gene-of-interest at a site between exisiting SARS virus genes. Small nonstructural gene products are not depicted schematically.

FIGURE 14: Schematic depicting representative examples of SARS virus replicons. Small nonstructural gene products are not depicted schematically.

FIGURE 15: SARS virus nsp2 proteinase (3CLp) and identification of catalytic and substrate sites.

FIGURE 16: alignment of SARS virus nsp2 proteinase (3CLp) with that of avian IBV, MHV, and BCoV. Residues in dotted boxes are key residues the substrate sites (F, Y & H); residues in solid boxes are catalytic cysteine (C) and histidine (H) residues.

FIGURE 17: Genome organization of SARS coronavirus. Replicase and structural regions are shown, along with the predicted products of cleavage within ORF1a and ORF1b. The position of the 5' RNA leader sequence (L), the 3' poly(A) tract and the ribosomal frame-shift consensus between ORF1a and ORF1b are also indicated. Each box represent a protein product. They are shaded according to the level of amino acid identity with corresponding proteins of other coronaviruses (see also Table 2). The SARS-specific genes are white. Positions of the 9 SARS-specific six-base IG sequences (5'-ACGAAC-3'; SEQ ID NO 7293) are indicated by arrows.

FIGURE 18: Genome organization of Coronaviruses representative of group 1 (HCoV-229E, accession number: AF304460), group 2 (mouse hepatitis virus MHV, accession number: NC_001846), group 3 (avian infectious bronchitis virus AIBV, accession number: NC_001451)

and SARS coronavirus. Other completely sequenced coronaviruses used in this study are available at the following accession numbers: porcine epidemic diarrhea virus (PEDV), AF353511; transmissible gastroenteritis virus (TGV), NC_002306; Bovine coronavirus (BCoV): AF220295. Red boxes represent group-specific genes. The position of the leader RNA sequence and poly(A) tract is also indicated in genomes where they are reported. The position of specific IG sequences is indicated by circles of different shades. In the SARS genome, we also find three IG sequences specific for group 2 coronavirus.

FIGURE 19: Topological model predicted for the spike protein anchored to the viral membrane. Structural and predicted functional domains are indicated. The N-terminal region (S1) is predicted to contain the receptor binding domain. Two coiled coil regions within the S2 domain, partially superimposed to leucine zipper motifs are presumably involved in oligomerization. The hydrophobic domain is responsible for membrane anchoring.

FIGURE 20: Phylogenetic tree obtained from the multiple sequence alignment of a 922 bp internal region of the *pol* gene from 12 coronaviruses and SARS. Numbers at the nodes represent the result of a bootstrap analysis and strongly support the branches. Sequences not available within the complete coronavirus genomes have been retrieved from GenBank at the following accession numbers: hemagglutinating encephalomyelitis virus of swine (PHEV), AF124988, Human OC43 virus (OC43), AF124989, canine coronavirus (CCV), AF124986, feline infectious peritonitis virus (FIPV), AF124987, turkey coronavirus (TCV), AF124991, syaloacryoadenitis virus of rats (SDAV), AF124990.

FIGURE 21: 21A. Unrooted tree obtained from the alignment of consensus sequences of the group I and group II S1 domain of spike proteins (G1_cons and G2_cons) with those of a group 3 spike (AIBV) and the spike of SARS virus. The number indicates the result of a bootstrap analysis. The sequences used to generate the consensus profile from group 1 are: HcoV-229E, accession number P15423; porcine epidemic diarrhea virus (PEDV), acc no: NP_598310; transmissible gastroenteritis virus (TGV), acc no: NP_058424; Canine coronavirus (CCV), acc no: S41453; porcine respiratory virus (PRV), acc no: S24284; feline infectious peritonitis virus (FIPV), acc no: VGIH79. The sequences used to generate the consensus profile from group 2 are: mouse hepatitis virus (MHV), acc no: NP_045300; Bovine coronavirus (BCoV), acc no: NP_150077; Human coronavirus OC43, acc no: P36334; hemagglutinating encephalomyelitis virus of swine (PHEV), acc no: AAL80031; for group 3, only the sequence of the spike protein of avian infectious bronchitis virus (AIBV), acc no: AAO34396 was used. 21B: Schematic representation of cysteine positions in S1 domains of group 1, 2 and 3, compared to the SARS spike. Horizontal bars represent the S1 amino acid sequences (in the case of SARS and AIBV) or the consensus profiles (generated from group 1, G1_cons, and from group 2, G2_cons). The

length of the bars are not to scale. Relative cysteine positions are indicated by rectangle bars. Only cysteines perfectly conserved within each consensus are reported. Lines connect cysteines conserved between the SARS S1 domain and the consensus sequences as shown.

FIGURE 22: illustration of a Neisseria Adhesin A protein (NadA).

FIGURE 23: Raw translation from SARS coronavirus genome (reading frame +1).

FIGURE 24: Raw translation from SARS coronavirus genome (reading frame +3)

FIGURE 25: 1b and Spike open reading frames, separated by *.

FIGURE 26: SARS growth in vero cells.

FIGURE 27: chromatogram of the capture step of SARS coronavirus on Matrix Cellufine Sulfate Superformance 150/10. Analysis was on 100ml coronoavirus harvest. The left Y axis shows absorbance at 280nm. The right Y axis shows the gradient (%B). The X axis shows the volume (ml).

FIGURE 28: Silver-stained MCS chromatography fractions. Lanes are: (1) marker; (2) coronavirus vero cell harvest; (3) coronavirus vero cell harvest, after 0.65μm filtration; (4) flowthrough; (5) wash; (6) 20% peak (virus peak). Lanes were loaded with 1 μg of test protein.

FIGURE 29: Western Blot of MCS chromatography fractions. Lanes are as described for Fig.28.

FIGURE 30: Linear density gradient ultracentrifugation, 15-60% sucrose (SW28, 2 hours, 20000 rpm). The graph shows protein concentration (■) and sucrose concentration (♦).

FIGURE 31: Silver-stained density gradient fractions on NuPage 4-12% Bis-Tris-Ge (Novex), reduced conditions, heated for 10 minutes at 70°C. Lanes are: (1) marker; (2) 20% peak MCS; (3) density gradient fraction 11; (4) density gradient fraction 12; (5) density gradient fraction 13; (6) density gradient fraction 14; (7) density gradient fraction 15; (8) density gradient fraction 16; (9) density gradient fraction 17. The bulk of proteins was in fractions 15 to 17. Lanes 2, 8 and 9 were loaded with 1μg protein.

FIGURE 32: Chromatogram of the Capture Step of SARS coronavirus on MCS. Details are as for Figure 27, except that 200ml harvest was used.

FIGURE 33: Silverstain (left) and Western Blot (right) of chromatographic fractions. Lanes are as described for Figures 28 and 29, except that lane (6) is the 5% peak. Treatment before SDS-PAGE was at room temperature for 30 minutes.

FIGURE 34: Density Gradient Ultracentrifugation, 15-40% sucrose (SW28, 2 hours, 20000 rpm). The graph shows protein concentration (■) and sucrose concentration (♦).

FIGURE 35: Silverstain (left) and Western Blot (right) of Density Gradient Ultracentrifugation fractions on NuPage 4-12% Bis-Tris-Ge (Novex), reduced conditions. Lanes are: (1) marker; (2) density gradient fraction 6; (3) density gradient fraction 7; (4) density gradient fraction 8; (5) density gradient fraction 9; (6) density gradient fraction 10; (7) density gradient fraction 15. Fractions 7-10 (lanes 3-6) contained pure coronavirus proteins. The bulk of impurities was in fraction 15 (lane 7). Lanes 2, 8 and 9 were loaded with ~1μg protein. Treatment before SDS-PAGE was at room temperature for 30 minutes.

FIGURE 36: EM pictures of Density Gradient Fractions 8-10. Figure 36A shows fraction 8; Figure 36B shows fraction 9; Figure 36C shows fraction 10.

FIGURE 37: Spike/NadA fusion constructs.

FIGURES 38 and 39: Results of the expression in *E.coli* of $S1_L$, $S1_L$-NadA and $S1_L$-$NadA_{\Delta anchor}$. Figure 38 shows SDS-PAGE analysis of total lysates from BL21(DE3)/pET, BL21(DE3)/pET-$S1_L$ and BL21(DE3)/pET-$S1_L$-$NadA_{\Delta anchor}$. The bands are indicated by an arrow, and the three lanes are, from left to right: BL21(DE3)/pET; BL21(DE3)/pET-$S1_L$; BL21(DE3)/pET-$S1_L$-$NadA_{\Delta anchor}$. Figure 39 shows (39A) SDS-PAGE and (39B) western blot analyses of total lysates from BL21(DE3)/pET, BL21(DE3)/pET-$S1_L$-NadA (grown under un-induced condition) and BL21(DE3)/pET-$S1_L$-NadA (grown under induced condition). The bands are indicated by an arrow, and lanes are, from left to right: BL21(DE3)/pET; BL21(DE3)/pET-$S1_L$-NadA; BL21(DE3)/pET-$S1_L$-NadA. The western blot shows the presence of oligomeric forms of the protein.

FIGURE 40: Schematic of SARS Spike clones.

FIGURE 41: Transient Expression of SARS Spike Proteins (western blot of COS7 cell lysate). Each lane of the 4-20% TG SDS gel was loaded with 20μg cell lysate (total 1.2mg). The labeling antibodies are shown.

FIGURE 42: Western blot analyses of COS7 cell lysates on 4% TG SDS gel showing oligomerization state of intracellular S molecules.

FIGURE 43: Western blot analyses of COS7 cell lysates on 4-20% TG SDS gel showing Transient Expression of SARS Spike Proteins. Lanes are: (1) mock, AF; (2) mock, DF; (3) nSh, AF; (4) nSh, DF; (5) nShΔTC, AF; (6) nShΔTC, DF. Each lane was loaded with 5μl of each sample, 400μl total. The blot was labeled with antibody against the His-tagged protein.

FIGURE 44: Western blot analyses of COS7 cell medium on 4-20% TG SDS gel showing Transient Expression of SARS Spike Proteins. Truncated spike protein is secreted. Spike proteins were purified from the culture medium (from a 10cm plate), first by a ConA column and then finally by His•tag Magnetic beads. Each lane was loaded with one third of the material.

FIGURE 45: Western blot analyses of COS7 cell lysates on 4-20% TG SDS gel showing glycoslation of SARS spike proteins. In the two left-hand blots (lanes 1-5), samples were boiled in SDS and β-mercaptoethanol; in the two right-hand blots (lanes 6-11), samples were in SDS only, with no boiling. Lanes 1-8 were labeled with a monoclonal raised against the His-tag protein; labes 9-11 were labeled with rabbit anti-SARS antibody.

FIGURE 46: Effect of SARS spike protein expression on cell viability.

FIGURE 47: Western blot analyses of COS7 cell lysates on 4% TG SDS gels showing oligomerization state of intracellular spike molecules. Blots were labeled with anti-His-tag mAb. The membrane fraction of COS7 cell lysate was fractionated by a sizing column before loading the lanes. Fractions 7 to 14 show bands with kDa values of: 71000, 1400, 898, 572, 365,232, 148 and 99, respectively.

FIGURE 48: Fractionation of cells into aqueous and detergent fractions.

FIGURE 49: Schematic of constructs for use in OMV preparation.

FIGURE 50: SARS HR1 and HR2 constructs.

FIGURE 51: Vaccine protection froms SARS in Balb/c mouse model.

FIGURE 52: Expressed on Spike protein in transfected 293 cell lysates (52A) or COS7 cell culture supernatants (52B). Proteins were separated on 4-20% TG SDS gels. The label was anti-His-tag, except for the right-hand three lanes of 52B, where the label was rabbit anti-SARS serum. In Figure 52A, the left-hand three lanes were treated with DTT and were boiled, but neither treatment was used for the right-hand three lanes. In Figure 52B, no DTT was used, but all lanes were heated to 80°C for 5 minutes.

FIGURE 53: Western blot of Spike proteins expressed in COS7 cells. Proteins were incubated at room temperature (RT), 80°C or 100°C to check for any effect on molecular weight. FIGURE 54 shows similar experiments on SARS virions.

FIGURE 55: Results of a pulse chase experiment, showing expression and processing of SARS spike protein following infection with alphavirus replicon particles. Cells were treated with or without EndoH as shown.

FIGURE 56: Effect of heating on Spike protein trimers.

FIGURE 57: Coomassie blue-stained gel of yeast-expressed proteins. Lanes are: 1-See Blue Standard (10μl); 2-pAB24 gbl (20μg); 3-SARS Spike S1 c.1 gbl (20μg); 4-SARS Spike S1 c.2 gbl (20μg); 5-See Blue Standard (10μl); 6-pAB24 ip (5μl); 7-SARS Spike S1 c.1 (5μl); 8-SARS Spike S1 c.2 (5μl).

FIGURES 58 to 64: Schematics of preparation of yeast expression constructs.

FIGURES 65 to 66: Yeast-expressed sequences for Spike.

FIGURE 67: Western blots showing expression of SARS spike protein from alphavirus replicon particles and replicon RNA. Figure 67A was run under non-reducing conditions and at room temperature (*i.e.* no heating), with lanes: (1) VEE/SIN-spike infection; (2) VEE/SIN-GFP infection; (3) Replicon-spike RNA transfection; (4) Replicon-GFP RNA transfection. Figure 67B was run with SARS virions at different temperatures, as shown.

FIGURE 68: induction of antibody responses in mice. Vaccine groups are: (1) Inactivated SARS Virus; (2) Truncated Recombinant Spike Protein; (3) Full length Spike: DNA+DNA.PLG+ Alphavirus; (4) Full length Spike: Alphavirus particles only.

FIGURE 69: Binding of human monoclonal antibody S3.2 to purified truncated Spike protein. The X-axis shows antibody concentration, and the Y-axis shows ELISA absorbance. The interpolation result is 2158.13.

FIGURE 70: Geometric mean ELISA titers of antibodies induced by the SARS-CoV spike protein delivered as different vaccines (left to right: inactivated virus; 3μg truncated spike protein; 75μg DNA encoding truncated spike protein.

FIGURE 71: Neutralization titers after immunization with (left) nSdΔTC protein or (right) DNA encoding nSdΔTC, delivered on PLG.

FIGURE 72: Correlation between the spike antigen binding and neutralizing antibodies

FIGURE 73: Western blot of CHO cell lines expressing Spike protein in full-length form (left) or in truncated form (right). Proteins were separated by 4-12% SDS-PAGE, with boiling in DTT and staining by polyclonal serum.

FIGURE 74: Structural components of SARS-CoV spike glycoprotein and expression construct. L denotes leader peptide (residues 1-13), TM the transmembrane, and Cy the cytoplasmic tail segments. The hexa-His tags are not shown.

FIGURE 75: Western blot analysis of SARS spike proteins expressed in COS7 cells. In Figure 75A, COS7 cells were transfected with indicated plasmid constructs and the expressed proteins in cell lysates 48 hr post-transfection were analysed by SDS-PAGE (4-20% polyacrylamide) in reducing and denaturing condtions, with proteins visualized by anti-histidine Mab. In Figure 75B, proteins were collected from cell culture medium 48 hr post-transfection and purified first by a ConA column and then by His-tag magnetic beads. Purifed proteins were analysed by SDS-PAGE (4-20% polyacrylamide) and were visualized by anti-SARS rabbit serum.

FIGURE 76: Endo H sensitivity of C-terminal truncated spike protein (SΔ) found in cell lysate (lanes 1,2) and culture medium (lanes 3,4). Positions of internal SΔ protein and secreted SΔ protein are marked with arrow heads.

FIGURE 77: Oligomeric status of the SARS spike protein. Recombinant S protein oligomer in COS7 cells transfected with the full-length spike construct (nSh). The cell lysates were treated with DTT and/or heat as indicated above each lane. The different forms of S protein in treated and untreated samples were visualized by SDS-PAGE (4% polyacrylamide) and Western blot analysis using anti-histidine MAb.

FIGURE 78: Effect of heat denaturation on the oligomeric status of recombinant S protein in the absence of DTT. The COS7 cell lysates were heated before the electrophoresis as indicated and the S proteins were visualized as described fogiFigure 77.

FIGURE 79: Effect of heat denaturation on the oligomeric status of spike protein in SARS virion particles. SARS-CoV were grown in Vero cells, purified and solubilized from the virion particles by SDS, heat-denatured as indicated and visualized as described in Figure 77, except that rabbit antiserum against the purified virus was used as a probe.

FIGURE 80: Analysis of the oligomeric status of SARS virion spike protein by cross-linking experiment. Solubilized SARS virion proteins were treated with DMS. Both untreated (–) and DMS treated (+) virion proteins were heat denatured in the absence of DTT and visualized by 4% PAGE followed by silver staining.

FIGURES 81 & 82: Analysis of the oligomeric status of truncated spike protein by heat denaturation. Truncated spike protein within COS7 cell lysates (81) or secreted into culture medium (82) were heat denatured as indicated in the absence of DTT and visualized by Western blot analysis.

FIGURE 83: Reactivity of deglycosylated full-length spike oligomer with conformational and non-conformational antibody. The full-length recombinant spike oligomer was partially deglycosylated with PNGase F in non denaturating condition and visualized by Western blot analysis using anti-histidine Mab (lane 1,2,3) or rabbit antiserum against purified SARS CoV (lane 4,5,6).

FIGURE 84: Localization of expressed SARS spike proteins in fractionated COS7 cell lysate visualized by western blot. Cells were transfected with indicated plasmids and lysed with Dounce homogeniser in hypotonic buffer 48 hr post transfection. Cell lysate was centrifuged to obtain soluble cytosol and insoluble membrane fraction that was further solublized by 4% Triton X-100. Proteins were heated with SDS at 80 C and analysed by SDS-PAGE (4-20% polyacrylamide) in reducing condtion. Proteins were visualized by anti-histidine Mab. The

cytosol fractions were loaded in lanes 1, 3, and 5 and the membrane fractions were loaded in lanes 2, 4, and 6.

FIGURE 85: Intracellular and surface expression of recombinant full-length (A,D) or truncated (B,E) spike protein in COS7 cells. The cells were fixed at 48 hrs posttransfection and either treated with detergent (Cytofix/perm, BD Biosciences) for intracellular immunofluorescence (A,B,C) or with 2% paraformaldehyde for cell surface immunofluorescence observation (D,E,F) at x40 magnification. Mock transfected cells (C,F) were included as controls.

FIGURES 86-105: SDS-PAGE od *E.coli* expressed proteins. Tot = total protein; Sol = soluble protein fraction. Labels are protein names (Tables 26-30).

FIGURE 106: Immunofluorescence after administration of vector encoding optimsed N antigen.

FIGURE 107: Immunofluorescence of (A) native and (B) codon-optimsed M sequences.

FIGURE 108: Immunofluorescence of (A) native and (B) codon optimsed E sequences.

FIGURES 109-111: Western blots of Vero cells using rabbit antibodies obtained after immunization with spike proteins expressed in *E.coli*.

FIGURE 112: Spike protein expression in 293 cells. Lanes: (M) Markers; (1) Mock transfected; (2,6) cells expressing nS protein, lysate; (3,7) cells expressing nSdTC protein, lysate; (4,8) cells expressing nS protein, supernatant; (5,9) (4) cells expressing nSdTC protein, supernatant. Staining antibody: (2 to 5) mouse serum obtained after DNA immunization; (6 to 9) rabbit serum obtained after immunization with whole killed virus.

FIGURE 113: Six reading frames of SEQ ID NO: 9968.

FIGURE 114: Six reading frames of SEQ ID NO: 10033.

FIGURE 115: Alignment of bovine coronavirus pol 1ab (top row; SEQ ID NO: 10068), avian infectious bronchitis pol 1ab (second row; SEQ ID NO: 10069), murine hepatitis virus pol 1ab (third row; SEQ ID NO: 10070), SEQ ID NO^{S}: 9997/9998 (fourth row) and a consensus sequence (bottom row; SEQ ID NO: 10071).

FIGURE 116: Schematic of coronavirus genome organization.

FIGURE 117: Schematic of coronavirus ORF1a/ORF1b gene products, including "*" region.

FIGURE 118: Alignment.

FIGURE 119: Alternative start codons within SEQ ID NO: 10080.

FIGURE 120: Six reading frames of SEQ ID NO: 10084.

FIGURE 121: Alignment of SEQ ID NO: 10033 and SEQ ID NO: 10084.

FIGURE 122: Reading frames in SEQ ID NO: 10084.

FIGURE 123: Start codon analysis for SEQ ID NO: 10084.

FIGURE 124: BLAST analysis of SEQ ID NO: 10210.

FIGURE 125: Epitope analysis of SEQ ID NO: 10210 by either (13A) Hopp & Woods or (13B) Kyte & Doolittle.

FIGURE 126: Reading frames in SEQ ID NO: 10299.

FIGURE 127: Reading frames in SEQ ID NO: 10505.

FIGURE 128: Reading frames in SEQ ID NO: 11563.

FIGURE 129: Reading frames in SEQ ID NO: 10033.

FIGURE 130: Alignment of SEQ ID NO: 9997 and SEQ ID NO: 10033.

FIGURE 131: Reading frames in SEQ ID NO: 10299.

FIGURE 132: Reading frames in SEQ ID NO: 10505.

FIGURE 133: Western Blot of SARS protease purification fractions.

FIGURE 134: Cleavage of DABCYL-EDANS (a fluorescent tagged peptide with a SARS protease cleavage site) by SARS protease at different concentrations. The graph shows activity/concentration correlations with no protease (♦), 0.95 uM protease (■) and 2.86 uM protease (•).

In the event of a discrepancy between a sequence in the sequence listing and a sequence in the drawings, the drawings should take precedence.

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, *e.g.*, *Remington's Pharmaceutical Sciences*, Mack Publishing Company, Easton, Pa., 19th Edition (1995); *Methods In Enzymology* (S. Colowick and N. Kaplan, eds., Academic Press, Inc.); and *Handbook of Experimental Immunology*, Vols. I-IV (D.M. Weir and C.C. Blackwell, eds., 1986, Blackwell Scientific Publications); Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Handbook of Surface and Colloidal Chemistry* (Birdi, K.S. ed., CRC Press, 1997); *Short Protocols in Molecular Biology*, 4th ed. (Ausubel *et al.* eds., 1999, John Wiley & Sons); *Molecular Biology Techniques: An Intensive Laboratory Course*, (Ream *et al.*, eds., 1998, Academic Press); *PCR (Introduction to Biotechniques Series)*, 2nd ed. (Newton & Graham eds., 1997, Springer Verlag); Peters and Dalrymple, *Fields Virology* (2d ed), Fields *et al.* (eds.), B.N. Raven Press, New York, NY.

All publications, patents and patent applications cited herein, are hereby incorporated by reference in their entireties.

Severe Acute Respiratory Syndrome (SARS) virus has recently been identified as a new viral species. The SARS viral species includes the following isolates.

- two virus isolates described in Peiris *et al.* "Coronavirus as a possible cause of severe acute respiratory syndrome" *Lancet* published online at *http://image.thelancet.com/extras/03art3477web.pdf* on April 8 2003, incorporated herein by reference in its entirety and the sequences deposited with GenBank at accession number AY268070.
- the isolates and viral sequences described in Drosten *et al.*, "Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome", New England Journal of Medicine, published online at *http://www.nejm.org* on April 10, 2003.
- the isolates and viral sequences described on the website of the WHO network on March 25 and 24, 2003.
- the isolates and viral sequences described in Tsang *et al.*, "A Cluster of Cases of Severe Acute Respiratory Syndrome in Hong Kong" *New England Journal of Medicine*, published online at *http://www.nejm.org* on March 31, 2003.
- the isolates and viral sequences described in Poutanen *et al.*, "Identification of Severe Acute Respiratory Syndrome in Canada" *New England Journal of Medicine*, published online at *http://www.nejm.org* on March 31, 2003.

As described in the *Lancet* article, a 646 base pair polynucleotide from the SARS virus has weak homology to viruses of the family *Cornoaviridae*. The *Lancet* article further reports that a deduced amino acid sequence (of 215 amino acids) from this sequence has about 57% sequence homology to the RNA polymerase of bovine coronavirus and murine hepatitis virus. Phylogenetic analysis of the protein sequences are also presented in the *Lancet* article showing that the polymerase sequence is most closely related to the group II coronaviruses.

Additional SARS viral isolates can be identified, isolated and/or sequenced by virologists skilled in the art. Virologists can readily identify new viral isolates as a SARS virus. Criteria which a virologist may use to identify new SARS isolates include: sequence homology of the new isolate to known SARS viral isolates; similar genomic organization of the new viral isolate to known SARS viral isolates; immunological (serologic) similarity or identity with known SARS viral isolates; pathology; and similarity of virion morphology with known SARS viral isolates; and similarity of infected cell morphology as that caused by known SARS viral isolates (visualized, for instance, by electron microscopy).

Methods for isolating and sequencing SARS viral isolates include the methods described by Peiris *et al.* in the *Lancet* paper. As reported in the *Lancet* paper, RNA from clinical samples

can be reverse transcribed with random hexamers and cDNA can be amplified with primers having sequences of SEQ ID NOS: 6584 & 6585 in the presence of 2.5 mmol/L magnesium chloride (94°C for 1 min, 50°C for 1 min, and 72°C for 1 min).

Reverse transcription of a viral isolate using random hexamers can be accomplished in an RT-PCR assay as follows. Virus isolates are propagated on mammalian cells, particularly fetal rhesus kidney cells. Total RNA from virus-infected and virus-uninfected fetal rhesus kidney cells is then isolated. RNA samples are reverse transcribed with a primer having SEQ ID NO: 6586. cDNA can be amplified by a primer having SEQ ID NO: 6587. Unique PCR products (in size) in the infected cell preparation are then cloned and sequenced, and genetic homology of the sequence compared with those in GenBank.

One skilled in the art would be able to identify and clone additional genomic regions using a variety of standard cloning techniques, such as, for example, using random primer RT-PCR and detection of sequences overlapping one or more of the above sequences, and/or using oligonucleotide primers, *e.g.*, degenerate primers, based on the sequences provided herein (see Figures 1-5, Figures 8-11, SEQ ID NOS: 3-20).

Cloning, sequencing and identification of SARS virus by one skilled in the art can be further facilitated by the use of polynucleotide sequences, particularly RNA polymerase sequences, from related Coronaviruses.

Sequence homology of new viral isolates with the known SARS isolates described above can be readily determined by one skilled in the art. New SARS isolates may be identified by a percent homology of viral nucleotide sequences of 99%, 95%, 92%, 90%, 85%, or 80% homology of the new virus to known SARS viral polynucleotide sequences. New SARS isolates may also be identified by percent homology of 99%, 95%, 92%, 90%, 85%, or 80% homology of the polypeptides encoded by the polynucleotides of the new virus and the polypeptides encoded by known SARS virus.

New SARS isolates may also be identified by a percent homology of 99%, 95%, 92%, 90%, 85%, or 80% homology of the polynucleotide sequence for specific genomic regions for the new virus with the polynucleotide sequence for specific genomic regions of the known SARS viruses. Additionally, new SARS isolates may be identified by a percent homology of 99%, 95%, 92%, 90%, 85%, or 80% homology of the polypeptide sequence encoded by the polynucleotide of specific genomic regions of the new SARS virus to the polypeptide sequence encoded by the polynucleotides of specific regions of the known SARS virus. These genomic regions may include regions (*e.g.*, gene products) which are typically in common among numerous coronaviruses, as well as group specific regions (*e.g.*, antigenic groups), such as, for example, any one of the following genomic regions which could be readily identified by a virologist skilled in the art: 5'untranslated region (UTR), leader sequence, ORF1a, ORF1b,

nonstructural protein 2 (NS2), hemagglutinin-esterase glycoprotein (HE) (also referred to as E3), spike glycoprotein (S) (also referred to as E2), ORF3a, ORF3b, ORF3x, nonstructural protein 4 (NS4), envelope (small membrane) protein (E) (also referred to as sM), membrane glycoprotein (M) (also referred to as E1), ORF5a, ORF5b, nucleocapsid phosphoprotein (N), ORF7a, ORF7b, intergenic sequences, 3'UTR, or RNA dependent RNA polymerase (pol). The SARS virus may have identifiable genomic regions with one or more the above-identified genomic regions. A SARS viral antigen includes a protein encoded by any one of these genomic regions. A SARS viral antigen may be a protein or a fragment thereof, which is highly conserved with coronaviruses. A SARS viral antigen may be a protein or fragment thereof, which is specific to the SARS virus (as compared to known cornaviruses). (See, Figures 1-5, Figures 8-11, SEQ ID NOS: 3-20).

One skilled in the art could also recognize electron microscopy of a SARS virus infected mammalian cell. Electron microscopy of SARS infected cells are shown in the *Lancet* paper. As discussed in the paper, electron microscopy of negative stained (3% potassium phospho-tungstate, pH 7.0) ultracentrifuged cell-culture extracts of SARS infected fetal rhesus kidney cells show the presence of pleomorphic enveloped virus particles of around 80-90 nm (range 70-130 nm) in diameter with surface morphology compatible with a coronavirus (see *Lancet* paper, Figure 1). Thin-section electron microscopy of infected cells reveals virus particles of 55-90 nm diameter within smooth walled vesicles in the cytoplasm (see *Lancet* paper, Figure 2B). Electron microscopy can also be used to observe virus particles at the cell surface. Electron microscopy of a human lung biopsy sample depicts similar viral morphology. See *Lancet* paper Figure 2A.

I. SARS POLYPEPTIDES AND POLYNUCLEOTIDES

The invention relates to nucleic acids and proteins from SARS virus. Such polynucleotides and polypeptides are exemplified further below.

In one embodiment, the polynucleotides of the invention do not include one of the following five primers, disclosed at *http://content.nejm.org/cgi/reprint/NEJMoa030781v2.pdf*: SEQ ID NOS: 6034-38.

The invention also includes polynucleotide sequences which can be used as probes for diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention includes a polynucleotide sequence comprising one or more of the primer sequences identified in SEQ ID NOS: 21-1020. The invention further includes polynucleotide sequence comprising the complement of one or more of the primer sequences identified in SEQ ID NOS: 21-1020.

The invention includes a polypeptide sequence comprising an amino acid sequence from the sequence shown in Figure 23. Such amino acid sequences are SEQ ID NOS: 6588-6809. The invention includes polypeptides comprising an amino acid sequence having sequence identity to these sequences, and the invention includes a fragment of a polypeptide comprising one of these sequences.

The invention includes a polypeptide comprising an amino acid sequence from the sequence shown in Figure 24. Such amino acid sequences are SEQ ID NOS: 6810-7179. The invention includes a protein comprising an amino acid sequence having sequence identity to these sequences, and the invention includes a fragment of a protein comprising one of these sequences.

The invention includes a protein comprising SEQ ID NO: 6039. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 6039. The invention includes a fragment of a polypeptide comprising SEQ ID NO: 6039. The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 6039, or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding SEQ ID NO: 6039, or a fragment thereof. The invention includes an immunogenic composition comprising a polypeptide comprising SEQ ID NO: 6039, or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising SEQ ID NO: 6039, or a fragment thereof. SEQ ID NO: 6039 demonstrates functional homology with ORF1a of coronaviruses.

Predicted transmembrane or hydrophobic regions of SEQ ID NO: 6039 are identified below. Although the polyprotein of coronaviruses is proteolytically cleaved into numerous smaller proteins, hydrophobic domains in the polyprotein are known to mediate the membrane association of the replication complex and to be able to dramatically alter the architecture of host cell membranes. Accordingly, the hydrophobic domains of the polyprotein are targets for genetic mutation to develop attenuated SARS virus vaccines. The hydrophobic domains are also targets for small molecule inhibitors of the SARS virus. The hydrophobic domains may also be used to generate antibodies specific to those regions to treat or prevent SARS virus infection.

Predicted Transmembrane Helices in SEQ ID NO: 6039

The sequence positions in brackets denominate the core region.
Only scores above 500 are considered significant.

Inside to outside helices : 43 found

from	to	score	center
100 (100)	118 (116)	103	107
473 (473)	488 (488)	1003	481
529 (532)	549 (549)	541	539
584 (584)	606 (601)	1049	594
773 (773)	791 (789)	514	782

Outside to inside helices : 43 found

from	to	score	center
94 (97)	118 (112)	291	104
400 (400)	418 (415)	243	407
473 (473)	488 (488)	1113	481
523 (528)	548 (548)	285	538
583 (583)	606 (601)	662	593

1071	(1071)1089	(1086)	243	1078
1121	(1121)1137	(1137)	459	1130
1679	(1679)1696	(1696)	404	1686
2098	(2102)2119	(2116)	509	2109
2145	(2145)2160	(2160)	797	2153
2206	(2209)2224	(2224)	2686	2216
2316	(2316)2332	(2332)	2123	2325
2335	(2339)2358	(2354)	2101	2346
2373	(2373)2390	(2390)	532	2380
2597	(2600)2615	(2615)	307	2607
2753	(2753)2770	(2768)	2242	2760
2831	(2833)2854	(2851)	759	2841
2879	(2882)2900	(2897)	526	2889
2990	(2996)3012	(3010)	1289	3003
3024	(3024)3042	(3039)	2281	3032
3054	(3058)3075	(3072)	2536	3065
3105	(3109)3127	(3123)	2010	3116
3143	(3143)3163	(3159)	184	3152
3253	(3255)3272	(3272)	319	3262
3346	(3346)3366	(3366)	203	3356
3375	(3375)3392	(3392)	305	3384
3438	(3438)3455	(3453)	1021	3445
3559	(3567)3584	(3581)	1885	3574
3589	(3589)3606	(3604)	2018	3596
3611	(3611)3629	(3629)	2304	3621
3659	(3659)3674	(3674)	1561	3667
3756	(3758)3777	(3774)	2352	3767
3890	(3890)3904	(3904)	485	3897
3916	(3919)3934	(3934)	241	3926
4035	(4035)4051	(4051)	335	4044
4217	(4217)4232	(4232)	272	4224
4239	(4239)4257	(4254)	402	4247

776	(776) 791	(791)	1435	783
1068	(1071)1089	(1086)	370	1078
1121	(1121)1137	(1137)	455	1130
1679	(1679)1696	(1694)	340	1686
2098	(2098)2119	(2116)	678	2109
2148	(2148)2163	(2163)	434	2155
2208	(2210)2231	(2226)	2389	2219
2309	(2309)2332	(2326)	1773	2318
2342	(2342)2368	(2360)	1666	2353
2373	(2373)2390	(2390)	254	2380
2753	(2755)2770	(2770)	2119	2763
2832	(2835)2854	(2851)	687	2844
2858	(2858)2873	(2873)	253	2866
2879	(2882)2899	(2899)	400	2889
2990	(2990)3005	(3005)	875	2998
3020	(3024)3042	(3042)	2795	3032
3059	(3059)3075	(3075)	2137	3067
3105	(3108)3127	(3123)	1902	3115
3142	(3145)3162	(3162)	540	3152
3343	(3351)3366	(3366)	496	3358
3437	(3437)3453	(3453)	848	3444
3489	(3491)3508	(3505)	302	3498
3560	(3560)3577	(3577)	1460	3569
3591	(3591)3606	(3606)	2193	3598
3610	(3610)3627	(3627)	1484	3620
3656	(3658)3678	(3675)	1240	3668
3681	(3684)3701	(3699)	590	3691
3710	(3713)3738	(3728)	1696	3721
3723	(3723)3738	(3738)	1670	3730
3760	(3760)3777	(3775)	2367	3767
3881	(3884)3902	(3900)	249	3892
4099	(4099)4114	(4114)	389	4106
4234	(4234)4254	(4249)	325	4241
4338	(4341)4360	(4360)	505	4348

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6039, wherein said fragment comprises an amino acid sequence including one or more of the hydrophobic transmembrane sequences identified above. The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6039 wherein said fragment comprises one or more of the following polypeptide sequences of SEQ ID NO: 6039: 473-488, 529-549, 584-606, 773-791, 2098-2119, 2145-2160, 2206-2224, 2316-2332, 2335-2358, 2373-2390, 2753-2770, 2831-2854, 2879-2900, 2990-3012, 3024-3042, 3054-3075, 3105-3127, 3438-3455, 3559-3584, 3589-3606, 3611-3629, 3659-3674, 3756-3777, 473-488, 583-606, 776-791, 2098-2119, 2208-2231, 2309-2332, 2342-2368, 2753-2770, 2832-2854, 2990-3005, 3020-3042, 3059-3075, 3105-3127, 3142-3162, 3437-3453, 3560-3577, 3591-3606, 3610-3627, 3656-3678, 3710-3738, 3723-3738, and 3760-3777. Preferably, the fragment comprises one or more of the following polypeptide sequences of SEQ ID NO: 6039: 2206-2224, 2316-2332, 2335-2358, 2753-2770, 3024-3042, 3054-3075, 3105-3127, 3589-3606, 3611-3629, 3756-3777, 2208-2231, 2753-2770, 3020-3042, 3059-3075, and 3591-3606. Preferably, the fragment comprises one or more of the following

polypeptide sequences of SEQ ID NO: 6039: 2206-2224 and 3020-3042. The invention also includes polynucleotides encoding each of the polypeptide fragments identified above.

The invention includes an attenuated SARS virus wherein said attenuated SARS virus contains an addition, deletion or substitution in the polynucleotides encoding for one of the hydrophobic domains identified above. The invention also includes a method for creating an attenuated SARS virus comprising mutating a SARS virus by adding, deleting or substituting the viral genome of the SARS virus to alter the coding of one or more of the hydrophobic domains of SEQ ID NO: 6039 identified above.

The invention includes an antibody which specifically identifies one or more of the hydrophobic regions of SEQ ID NO: 6039 identified above. The invention includes a small molecule which binds to, interferes with the hydrophobicity of or otherwise disrupts one or more of the hydrophobic regions of SEQ ID NO: 6039 identified above.

Predicted N-glycosylation sites of SEQ ID NO: 6039 are identified in the chart below.

Prediction of N-glycosylation sites in SEQ ID NO: 6039

Position		Potential	Jury agreement	NGlyc result	
48	NGTC SEQ ID NO: 7180	0.6371	(7/9)	+	
389	NHSN SEQ ID NO: 7181	0.6132	(6/9)	+	
916	NFSS SEQ ID NO: 7182	0.5807	(7/9)	+	
1628	NHTK SEQ ID NO: 7183	0.5610	(7/9)	+	
1696	NKTV SEQ ID NO: 7184	0.5297	(5/9)	+	
2031	NPTI SEQ ID NO: 9764	0.5299	(5/9)	+	WARNING: PRO-X1.
2249	NSSN SEQ ID NO: 7185	0.6329	(9/9)	++	
2459	NPTD SEQ ID NO: 9765	0.5599	(6/9)	+	WARNING: PRO-X1.
2685	NVSL SEQ ID NO: 7186	0.6071	(8/9)	+	
4233	NATE SEQ ID NO: 7187	0.6144	(7/9)	+	

Accordingly, the invention comprises a fragment of SEQ ID NO: 6039 wherein said fragment comprises an amino acid sequence which includes one or more of the N-glycosylation sites identified above. Preferably, the fragment comprises one or more sequences selected from the group consisting of SEQ ID NOS: 7180-7187 & 9764-9765. Preferably, the fragment comprises the amino acid sequence NSSN (SEQ ID NO: 7185).

The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6039 wherein said fragment does not include one or more of the glycosylation sites identified above. The invention also includes a polynucleotide encoding such a polypeptide.

T-epitopes for SEQ ID NO: 6039 are identified in Table 13. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified as SEQ ID NOS: 7400-7639; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the

polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified as SEQ ID NOS: 7400-7639, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus.

The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The ORF1a and ORF1b sequences of coronaviruses are typically translated as a single ORF1ab polyprotein. Slippage of the ribosome during translation generates an a-1 frameshift. One region of such slippage is illustrated below:

```
   gggttttacacttagaaacacagtctgtaccgtctgcggaatgtggaaaggttatggctgtagttgtga
+1   G  F  T  L  R  N  T  V  C  T  V  C  G  M  W  K  G  Y  G  C  S  C  D
+3  G  F  Y  T  -  K  H  S  L  Y  R  L  R  N  V  E  R  L  W  L  -  L  -
   ccaactccgcgaacccttgatgcagtctgcggatgcatcaacgttttttaaacgggtttgcggtgtaagt
+1   Q  L  R  E  P  L  M  Q  S  A  D  A  S  T  F  L  N  G  F  A  V  -  V
+3  P  T  P  R  T  L  D  A  V  C  G  C  I  N  V  F  K  R  V  C  G  V  S
   gcagcccgtcttacaccgtgcggcacaggcactagtactg  (SEQ ID NO: 7224)
+1   Q  P  V  L  H  R  A  A  Q  A  L  V  L   (SEQ ID NOS: 7225-7226)
+3  A  A  R  L  T  P  C  G  T  G  T  S  T    (SEQ ID NOS: 7227-7229)
```

which would generate the following translational slippage (SEQ ID NOS: 7230-7231):

```
ccaactccgcgaacccttgatgcagtctgcggatgcatcaacgttttttaaacgggtttgcggtgtaagt
 Q  L  R  E  P  L  M  Q  S  A  D  A  S  T  F  L  N  R  V  C  G  V  S
```

Accordingly, the invention includes a polypeptide comprising SEQ ID NO: 7232. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 7232. The invention includes a fragment of a polypeptide comprising SEQ ID NO: 7232 The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 7232 or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding SEQ ID NO: 7232 or a fragment thereof. The invention includes an immunogenic composition comprising a polypeptide comprising SEQ ID NO: 7232 or a

fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising SEQ ID NO: 7232 or a fragment thereof.

The invention also includes a polypeptide comprising amino acid sequence X_1-X_2-X_3, where X_1 is SEQ ID NO: 7233, X_2 is from one to ten amino acids, and X_3 is SEQ ID NO: 7234. X_2 can comprise any sequence of one to ten amino acids (SEQ ID NOS: 7235-7244) but, in preferred embodiments, X_2 is selected from the group consisting of F, FL, FLN, FLNR (SEQ ID NO: 7245), FLNRV (SEQ ID NO: 7246) and FLNRVC (SEQ ID NO: 7247). Preferably, X_2 is SEQ ID NO: 7247. These preferred embodiments are shown as SEQ ID NOS: 7248-7253.

The invention includes a polypeptide comprising an amino acid sequence having sequence identity to said amino acid sequences X_1-X_2-X_3. The invention includes a fragment of a polypeptide comprising said amino acid sequences X_1-X_2-X_3. The invention includes a diagnostic kit comprising a polypeptide comprising said amino acid sequences X_1-X_2-X_3 or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding said amino acid sequences X_1-X_2-X_3 or a fragment thereof. The invention includes an immunogenic composition comprising a polypeptide comprising said amino acid sequences X_1-X_2-X_3 or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising said amino acid sequences X_1-X_2-X_3 or a fragment thereof.

The amino acid sequences X_1-X_2-X_3 (*i.e.* SEQ ID NOS: 7235-7244) demonstrate functional homology with the polyprotein of murine hepatitis virus. This polyprotein is cleaved to produce multiple proteins. Proteins which can be generated from the X_1-X_2-X_3 polyprotein, where X_2 is six amino acids (SEQ ID NO: 7240) are listed below.

Mouse virus protein	Coordinates in Mouse virus	Coordinates in SEQ ID NO: 7240
Nsp2	3334-3636	3241-3546
Nsp3	3637-3923	3547-3836
Nsp4	3924-4015 (or 4012)	3837-3919
Nsp5	4016 (or 4013)-4209	3920-4117
Nsp6	4210-4319	4118-4230
Nsp7	4320-4456	4231-4369
Nsp9	4457-5384	4370-5301
Nsp10	5385-5984	5302-5902
Nsp11	5985-6505	5903-6429
Nsp12	6506-6879	6430-6775
Nsp13	6880-7178	6776-7073

The invention includes a fragment of the amino acid sequence X_1-X_2-X_3 (*i.e.* SEQ ID NOS: 7235-7244) wherein the fragment comprises one of the polypeptide sequences identified in the above table. The invention further includes a fragment of the amino acid sequence X_1-X_2-X_3 wherein said fragment comprises a polypeptide sequence which has a serine at its N-terminus and a glutamine at its C-terminus. The invention further includes a fragment of the amino acid sequence X_1-X_2-X_3 wherein said fragment comprises a polypeptide sequence which has an

Alanine at its N-terminus and a glutamine at its C-terminus. The invention further includes a fragment of the amino acid sequence X_1-X_2-X_3 wherein said fragment comprises a polypeptide sequence which has a Asparagine at its N-terminus and a glutamine at its C-terminus. The invention further includes a fragment of the amino acid sequence X_1-X_2-X_3 wherein said fragment comprises a Cysteine at its N-terminus and a Glutamine at its C-terminus. Each of the fragments identified above can be used in fusion proteins.

The invention includes a diagnostic kit comprising a polypeptide comprising at least one of the fragments of the amino acid sequence X_1-X_2-X_3 (*i.e.* SEQ ID NOS: 7235-7244) identified in the above paragraph. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding at least one of the fragments of the amino acid sequence X_1-X_2-X_3 identified in the above paragraph. The invention includes an immunogenic composition comprising a polypeptide comprising at least one of the fragments of the amino acid sequence X_1-X_2-X_3 identified in the above paragraph. The invention includes an antibody which recognizes a polypeptide comprising at least one of the fragments of the amino acid sequence X_1-X_2-X_3 identified in the above paragraph.

Predicted N-glycosylation sites of the amino acid sequence X_1-X_2-X_3 when X_2 is six amino acids are identified at the asparagines located at the following amino acid positions 48; 389; 556; 916; 1628; 1696; 1899; 2079; 2249; 2252; 2507; 2685; 3303; 3373; 3382; 3720; 4150; 4233; 4240; 5016; 5280; 5403; 5558; 5650; 5905; 6031; 6130; 6474; 6918; 6973. Accordingly, the invention comprises a fragment of SEQ ID NO: 7239 wherein said fragment is at least ten amino acids and wherein said fragment comprises one or more of the asparagines from the amino acid positions of SEQ ID NO: 7239 selected from the group consisting of 8; 389; 556; 916; 1628; 1696; 1899; 2079; 2249; 2252; 2507; 2685; 3303; 3373; 3382; 3720; 4150; 4233; 4240; 5016; 5280; 5403; 5558; 5650; 5905; 6031; 6130; 6474; 6918; and 6973.

A zinc binding region 2 site within SEQ ID NOS: 7235-7244 is identified at amino acid residues 2102-2112 (SEQ ID NO: 7254 HGIAAINSVPW). The polypeptide of SEQ ID NOS: 7235-7244 will be processed by the SARS virus into multiple peptides. This zinc binding region falls within the nsp1 region of the polypeptide. SEQ ID NO: 7254 is a target for screening of chemical inhibitors to the SARS virus. The invention includes a polypeptide comprises SEQ ID NO: 7254. The invention includes a polynucleotide encoding the polypeptide sequence of SEQ ID NO: 7254. The invention includes a method of screening SEQ ID NO: 7254 for an inhibitor. The invention includes the recombinant expression of SEQ ID NO: 7254 in a host cell. The invention includes a fragment of SEQ ID NOS: 7235-7244, wherein said fragment comprises SEQ ID NO: 7254. The invention includes a polypeptide comprising SEQ ID NO: 7254 wherein said polypeptide is complexed with a zinc ion. The invention includes a small molecule which

prevents a zinc ion from complexing with the polypeptide of SEQ ID NO: 7254. The invention includes a fusion protein wherein said fusion protein comprises SEQ ID NO: 7254.

The polyprotein encoded by the SARS virus will contain at least two protease domains: a papain-like cystein protease (PLP) and a chymotrypsin-picornavirus 3C-like protease (3CLp). (There may be more than one copy of the PLP domain). These proteases function to cleave the polyprotein into multiple smaller proteins. The 3C-like protease, also known as the "main protease" or Mpro, is itself cleaved from the polyprotein by its own autoprotease activity. See generally, Chapter 35 of *Fields Virology* (2nd ed), Fields *et al.* (eds.), B.N. Raven Press, New York, NY, and Anand *et al.*, *EMBO Journal* (2002) 21 (13): 3213-3224. This 3CLp generally corresponds with the Nsp2 region identified above.

The SARS virus 3CLp protein is further characterized by SEQ ID NO: 6569 (also SEQ ID NO: 9769), as shown in FIGURE 15.

FIGURE 16 also illustrates the SARS virus 3CLp, in allignment with the 3CLp of avian infectious bronchitis (IBV; SEQ ID NO: 6570), mouse hepatitis virus (MHV; SEQ ID NO: 6571), and bovine coronavirus (BCoV; SEQ ID NO: 6572). Accordingly, the invention includes a polypeptide sequence comprising SEQ ID NO: 6569, or a fragment thereof, or a polypeptide sequence having sequence identity thereto. The invention further includes a polynucleotide sequence encoding SEQ ID NO: 6569, or a fragment thereof. The invention includes a polynucleotide sequence encoding a polypeptide sequence having sequence identity to SEQ ID NO: 6569.

The invention further includes a method of screening for an inhibitor of the SARS virus 3CLp protein. In one embodiment, the invention includes a method of screening for an inhibitor of SEQ ID NO: 6569. The invention includes a method of recombinantly expressing the SARS virus 3CLp protein in a host cell. The invention includes a method of recombinantly expressing a polypeptide sequence comprising SEQ ID NO: 6569 or an enzymatically active fragment thereof or a polypeptide sequence having sequence identity thereto. The invention includes a small molecule which inhibits or reduces the proteolytic activity of the SARS virus 3CLp protein. The invention includes a small molecule which inhibits or reduced the proteolytic activity of the polypeptide comprising SEQ ID NO: 6569.

Catalytic residues of the SARS virus 3CLp are identified in FIGURE 15 and 16. Specifically, a catalytic histidine and a catalytic cysteine are identified. Such catalytic sites are targets for small molecules which could inhibit or reduce the protease activity of 3CLp. Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6569, wherein said fragment comprises at least one catalytic site. Preferably, the catalytic site is selected from the group consisting of the indicated catalytic histidine and the catalytic cysteine in FIGURE 15 and 16. The invention includes a polynucleotide encoding a polypeptide, wherein

said polypeptide comprises a fragment of SEQ ID NO: 6569, wherein said fragment comprises at least one catalytic site. Preferably, the catalytic site is selected from the group consisting of the indicated catalytic histidine and the catalytic cysteine.

The invention further includes a method of screening a compound library to identify a small molecule which inhbits a catalytic site of a SARS virus 3CLp. Preferably, the 3CLp comprises SEQ ID NO: 6569, or a fragment thereof, or a sequence having sequence identity thereto. The catalytic site is preferably selected from the group consisting of the indicated catalytic histidine and the catalytic cysteine in FIGURE 15 and 16.

The invention includes a small molecule which inhibits the catalytic site of a SARS virus 3CLp. Preferably, the 3CLp comprises SEQ ID NO: 6569, or a fragment thereof, or a sequence having sequence identity thereto. The catalytic site is preferably selected from the group consisting of the indicated catalytic histidine and the catalytic cysteine in FIGURE 15 and 16.

Residues of the substrate site of the SARS virus 3CLp are identified in FIGURE 15 and 16. Specifically, a substrate site is indicated at a phenylalanine, a tyrosine and a histidine. Such substrate sites are targets for small molecules which could inhibit or reduce the protease activity of 3CLp. Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6569, wherein said fragment comprises at least one substrate site. Preferably, the substrate site is selected from the group consisting of the indicated substrate phenylalanine, tyrosine and histidine in FIGURE 15 and 16. The invention includes a polynucleotide encoding a polypeptide, wherein said polypeptide comprises a fragment of SEQ ID NO: 6569, wherein said fragment comprises at least one substrate site. Preferably, the substrate site is selected from the group consisting of the indicated substrate phenylalanine, tyrosine and histidine in FIGURE 15 and 16.

The invention further includes a method of screening a compound library to identify a small molecule which blocks a substrate site of a SARS virus 3CLp. Preferably, the 3CLp comprises SEQ ID NO: 6569, or a fragment thereof, or a sequence having sequence identity thereto. The substrate site is preferably selected from the group consisting of the indicated substrate phenylalanine, tyrosine and histidine in FIGURE 15 and 16.

The invention includes a small molecule which inhibits the substrate site of a SARS virus 3CLp. Preferably, the 3CLp comprises SEQ ID NO: 6569, or a fragment thereof, or a sequence having sequence identity thereto. The substrate site is preferably selected from the group consisting of the indicated substrate phenylalanine, tyrosine and histidine in FIGURE 15 and 16.

The invention further includes a diagnostic kit comprising a polynucleotide encoding a SARS virus 3CLp or a fragment thereof. Preferably, the SARS virus 3CLp comprising SEQ ID NO: 6569 or a fragment thereof or a polypeptide sequence having sequence identity thereto. Preferably, the fragment comprising one or more sites selected from the group consisting of a

catalytic site and a substrate site. Preferably, the catalytic site is selected from the group consisting of one or more of the sites identified in FIGURE 15 and 16. Preferably, the substrate site is selected from the group consisting of one or more of the sites identified in FIGURE 15 and 16.

The invention further comprises a diagnostic kit comprising an antibody specific to a SARS virus 3CLp or a fragment thereof. Preferably, the antibody is specific to the polypeptide comprising SEQ ID NO: 6569 or a fragment thereof or a polypeptide sequence having sequence identity thereto. Preferably, the antibody is specific to one or more sites of a SARS virus 3CLp selected from the group consisting of a catalytic site and a substrate site. Preferably, the catalytic site is selected from the group consisting of one or more of the sites identified in FIGURE 15 and 16. Preferably, the substrate site is selected from the group consisting of one or more of the sites identified in FIGURE 15 and 16.

The invention includes a polypeptide comprising an amino acid sequence from the sequence shown in Figure 25. The two amino acid sequences within Figure 25, separated by a *, are SEQ ID NOS: 7188 & 7189. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to the figure 25 translation. The invention includes a fragment of a polypeptide comprising the figure 25 sequence. The invention includes a diagnostic kit comprising a polypeptide comprising the figure 25 translation, or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding the figure 25 translation, or a fragment thereof. The invention includes an immunogenic composition comprising a polypeptide comprising the figure 25 translation, or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising the figure 25 sequence, or a fragment thereof. The figure 25 sequence demonstrates functional homology with ORF1b of coronaviruses.

SEQ ID NO: 7188 is an open reading frame within Figure 25. The invention includes a polypeptide comprising SEQ ID NO: 7188. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 7188. The invention includes a fragment of a polypeptide comprising SEQ ID NO: 7188. The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 7188, or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding SEQ ID NO: 7188, or a fragment thereof. The invention includes an immunogenic composition comprising a polypeptide comprising SEQ ID NO: 7188, or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising SEQ ID NO: 7188, or a fragment thereof.

SEQ ID NO: 7190 is an open reading frame within SEQ ID NO: 7188. The invention includes a polypeptide comprising SEQ ID NO: 7190, a fragment thereof or a polypeptide having sequence identity thereto. The invention further includes a polynucleotide encoding SEQ

ID NO: 7190, a fragment thereof or a polypeptide sequence having sequence identity thereto. An example of a polynucleotide encoding SEQ ID NO: 7190 is given as SEQ ID NO: 7191.

SEQ ID NO: 7188 also contains an open reading frame comprising SEQ ID NO: 6042. The invention includes a polypeptide comprising SEQ ID NO: 6042. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 6042. The invention includes a fragment of a polypeptide comprising SEQ ID NO: 6042. The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 6042, or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding SEQ ID NO: 6042, or a fragment thereof. The invention includes an immunogenic composition comprising a polypeptide comprising SEQ ID NO: 6042, or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising SEQ ID NO: 6042, or a fragment thereof. SEQ ID NO: 6042 demonstrates functional homology to a coronavirus spike protein.

Predicted transmembrane regions of SEQ ID NO: 6042 are identified below.

Predicted Transmembrane helices of SEQ ID NO: 6042

The sequence positions in brackets denominate the core region.
Only scores above 500 are considered significant.

Inside to outside helices : 18 found

	from		to	score	center
1	(1)	16	(16)	959	9
233	(237)	257	(252)	905	244
345	(347)	364	(361)	490	354
345	(354)	369	(369)	420	362
497	(497)	513	(513)	239	506
573	(573)	588	(588)	811	580
645	(648)	666	(663)	302	656
690	(696)	714	(711)	428	704
857	(860)	882	(874)	1508	867
1031	(1031)	1046	(1046)	446	1039
1199	(1203)	1219	(1217)	2667	1210

Outside to inside helices : 13 found

	from		to	score	center
1	(1)	17	(17)	684	10
222	(222)	240	(237)	238	229
244	(247)	264	(264)	613	254
349	(355)	369	(369)	314	362
496	(496)	511	(511)	488	503
573	(573)	591	(591)	712	581
650	(652)	666	(666)	474	659
674	(679)	702	(696)	190	686
691	(696)	713	(711)	210	704
866	(868)	886	(886)	1172	876
1198	(1201)	1215	(1215)	3221	1208

SEQ ID NO: 6042, the spike protein, is a surface exposed polypeptide. Recombinant expression of a protein can be hindered by hydrophobic transmembrane regions. Accordingly, the invention includes a polypeptide comprising SEQ ID NO: 6042 wherein one or more of the hydrophobic regions identified above is removed. The invention further includes a polynucleotide encoding such a polypeptide. The invention includes recombinantly expressing the protein in a host cell. Primers for amplifying the gene for spike protein and fragments thereof, such as fragments encoding the soluble ectodomain, include SEQ ID NOS: 9753-9763 (Xiao *et al.* (2003) *Biochem Biophys Res Comm* 312:1159-1164).

Further characterization of SEQ ID NO: 6042 is set forth below.

PSORT --- Prediction of Protein Localization Sites

```
version 6.4(WWW)
SEQ ID NO: 6042 - 1255 Residues
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 2
     Position of the most N-terminal TMS: 496 at i=2
MTOP: membrane topology (Hartmann et al.)
     I(middle): 503   Charge diffirence(C-N):  1.0
McG: Examining signal sequence (McGeoch)
     Length of UR:   13
     Peak Value of UR:   3.28
     Net Charge of CR: 0
     Discriminant Score:      8.66
GvH: Examining signal sequence (von Heijne)
     Signal Score (-3.5): 5.94
     Possible cleavage site: 13
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 14
ALOM new cnt: 1 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
ALOM: finding transmembrane regions (Klein et al.)
     count: 1  value: -12.26  threshold: -2.0
     INTEGRAL    Likelihood =-12.26   Transmembrane 1202-1218 (1194-1228)
     PERIPHERAL  Likelihood =  0.16
     modified ALOM score:   2.55
>>> Seems to be a Type Ia membrane protein
     The cytoplasmic tail is from 1219 to 1255 (37 Residues)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway
(14) or uncleavable?
Gavel: Examining the boundary of mitochondrial targeting seq.
      motif at: 14
     Uncleavable? Ipos set to: 24
Discrimination of mitochondrial target seq.:
     positive ( 2.18)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway

*** Reasoning Step: 2

KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
  (Gavel position 24)  from: 1  to: 10  Score:  8.0
SKL motif (signal for peroxisomal protein):
     pos: 964(1255), count: 1   SRL
     SKL score (peroxisome):  0.1
Amino Acid Composition Tendency for Peroxisome:   1.37
     AAC not from the N-term., score modified
Peroxisomal proteins?   Status: notclr
     AAC score (peroxisome): 0.079
Amino Acid Composition tendency for lysosomal proteins
     score:  0.39  Status: notclr
GY motif in the tail of typeIa? (lysosomal)
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
```

```
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
Checking the RNA binding motif (nucleus or cytoplasm)
Nuclear Signal   Status: negative ( 0.00)
Type Ia is favored for plasma memb. proteins
Checking the NPXY motif..
Checking the YXRF motif..
Checking N-myristoylation..

----- Final Results -----
plasma membrane --- Certainty= 0.460(Affirmative) < succ>
microbody (peroxisome) --- Certainty= 0.171(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty= 0.100(Affirmative) < succ>
endoplasmic reticulum (lumen) --- Certainty= 0.100(Affirmative) < succ>
```

SEQ ID NO: 6042 appears to have a N-terminus signaling region, followed by a surface exposed region, followed by a transmembrane region followed by a C-terminus cytoplasmic domain region. Accordingly, the invention includes an immunogenic, surface exposed fragment of SEQ ID NO: 6042. Preferably, said fragment comprises an amino acid sequence which does not include the last 50 amino acids of the C-terminus of SEQ ID NO: 6042. Preferably, said fragment comprises an amino acid sequence which does not include the last 70 amino acids of the C-terminus of SEQ ID NO: 6042. Preferably, said fragment does not include a transdomain region of SEQ ID NO: 6042. Preferably, said fragment does not include a C-terminus cytoplasmic domain of SEQ ID NO: 6042. Preferably, said fragment does not include a N-terminus signal sequence. Preferably, said fragment does not include amino acids 1-10 of the N-terminus of SEQ ID NO: 6042. Preferably, said fragment does not include amino acids 1-14 of the N-terminus of SEQ ID NO: 6042. Two oligopeptide fragments of SEQ ID NO: 6042 that are able to elicit anti-spike antibodies are SEQ ID NOS: 7398 & 7399, as described (with additional C-terminus cysteines) by Xiao *et al.* (2003) *Biochem Biophys Res Comm* 312:1159-1164. C-terminal truncations of spike protein, with removal of part of the cytoplasmic region, or removal upto and including the transmembrane region, are described by Yang *et al.* (2004) *Nature* 428:561-564.

A variant of SEQ ID NO: 6042 that is included within the invention is SEQ ID NO: 9962. Compared to SEQ ID NO: 6042, this sequence has Ser at residue 581 instead of Ala, and has Phe at residue 1152 instead of Leu.

The spike protein of coronaviruses may be cleaved into two separate chains into S1 and S2. The chains may remain associated together to form a dimer or a trimer. Accordingly, the invention includes a polypeptide comprising SEQ ID NO: 6042 wherein said polypeptide has been cleaved into S1 and S2 domains. The invention further includes a polypeptide comprising SEQ ID NO: 6042 wherein amino acids 1-10, preferably amino acids 1-14 of the N-terminus are removed and further wherein SEQ ID NO: 6042 is cleaved into S1 and S2 domains. Preferably the polypeptide is in the form of a trimer.

The spike protein appears to form an alpha-helical structure in the transmembrane region of the protein, preferably in the S2 domain. This alpha-helical structure is thought to associate with at least two additional spike proteins to form a trimer. Helical or coiled regions of the spike protein are identified below. Predicted coiled-coils of SEQ ID NO: 6042 (spike protein) are at amino acids 900-1005 and 1151-1185 (see Figure 12).

Accordingly, the invention comprises a polypeptide sequence comprising a fragment of SEQ ID NO: 6042 wherein said fragment includes a coiled region of SEQ ID NO: 6042. Said fragment preferably includes the amino acid sequences selected from the group consisting of amino acid positions 900 to 1005 and amino acid positions 1151 to 1185 of SEQ ID NO: 6042. The invention comprises a polypeptide sequence comprising a fragment of SEQ ID NO: 6042, wherein said fragment does not include a coiled region of SEQ ID NO: 6042. Said fragment preferably includes the amino acid sequences selected from the group consisting of amino acid positions 900 to 1005 and amino acid positions 1151 and 1185 of SEQ ID NO: 6042.

The spike protein is believed to play an integral role in fusion and infection of Coronaviruses with mammalian host cells. Analysis of coronavirus spike proteins as well as similar surface proteins in other viruses has identified at least two structural motifs, typically located within the S2 domain, associated with this fusion event: heptad repeats (HR) and membrane fusion peptides.

At least two 4,3 hydrophobic heptad repeat (HR) domains are typically found in the ectodomain of the S2 domain of Coronaviruses. One heptad repeat region (HR1) is typically located adjacent to a fusion peptide while a second heptad region (HR2) is typically located near the C-terminus of the S2 domain, close to the transmembrane anchor. Heptad repeats are characteristic of coiled-coil structures and the heptad repeats found in viral surface proteins (such as coronavirus spike protein) are thought to form bundled helix structures which are involved in viral entry. *See* Bosch *et al.*, *J. Virology* (2003) 77:8801-8811 (Figure 1B of this reference illustrates an alignment of the HR1 and HR2 regions of five coronaviruses along with SARS, annotated "HCov-SARS").

Heptad repeats generally contain a repeating structure of seven amino acids, designated *a-b-c-d-e-f-g*, where hydrophobic sidechains of residues *a* and *d* typically form an apolar stripe, and electrostatic interactions are found in residues *e* and *g*. Position *a* is most frequently Leu, Ile or Ala and position *d* is usually Leu or Ala. Residues *e* and *g* are often Glu or Gln, with Arg and Lys also prominent at position *g*. Charged residues are common to positions *b*, *c* and *f* as these residues may be in contact with solvent. Exceptions to these general parameters are known. For instance Pro residues are sometimes found within the heptad.

The HR1 and HR2 sequences of an MHV strain have been postulated to assemble into a thermostable, oligomeric, alphahelical rold-like complex, with the HR1 and HR2 helices

oriented in an antiparallel manner. *Id.* In this same study, HR2 was asserted to be a strong inhitibor of both virus entry into the cell and cell-cell fusion.

HR1 and HR2 sequences have been identified in the SARS virus genome. The SARS virus HR1 region comprises approximately amino acids 879 to 1005 of SEQ ID NO: 6042 or fragments thereof capable of forming at least one alpha-helical turn. Preferably, said fragments comprise at least 7 (*e.g.*, at least 14, 21, 28, 35, 42, 49 or 56) amino acid residues. SEQ ID NO: 7192, includes amino acids 879 to 1005 of SEQ ID NO: 6042.

A preferred fragment of HR1 comprises amino acid residues 879 to 980 of SEQ ID NO: 6042. This preferred fragment is SEQ ID NO: 7193.

Another preferred fragment of HR1 comprises amino acid residues 901 to 1005 of SEQ ID NO: 6042. This preferred fragment is SEQ ID NO: 7194.

The SARS virus HR2 region comprises approximately amino acids 1144 to 1201 of SEQ ID NO: 6042, or fragments thereof capable of forming at least one alpha-helical turn. Preferably, said fragments comprise at least 7 (*e.g.*, at least 14, 21, 28, 35, 42, 49 or 56) amino acid residues. SEQ ID NO: 7195 includes amino acids 1144 to 1201. A preferred fragment of HR2 comprises amino acids 1144 to 1195 of SEQ ID NO: 6042. This preferred fragment is SEQ ID NO: 7196.

Membrane Fusion peptides sequences within the spike protein are also believed to participate in fusion (and infection) of the virus with a host cell. Fusion peptides generally comprise about 16 to 26 amino acid residues which are conserved within viral families. These Membrane Fusion peptides are relatively hydrophobic and generally show an asymmetric distribution of hydrophobitiy when modeled into an alpha helix. They are also generally rich in alanine and glycine.

At least three hydrophobic Membrane Fusion peptide regions have been identified within coronaviruses (PEP1, PEP2, and PEP3). *See,* Luo *et al.*, "Roles in Cell-Cell Fusion of Two Conserved Hydrophobic Regions in the Murine Coronavirus Spike Protein", Virology (1998) 244:483-494. Figure 1 of this paper shows an alignment of Membrane Fusion peptide sequences of Mouse Hepatitis Viris, Bovine Corona Virus, Feline Infectious Peritonitis Virus, Transmissible Gastroenteritis Virus and Infectious Bronchitis Virus. *See also*, Bosch *et al.*, "The Coronavirus Spike Protein is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex" Journal of Virology (2003) 77(16):8801-8811.

PEP1 (SEQ ID NO: 7197), PEP2 (SEQ ID NO: 7198) and PEP3 (SEQ ID NO: 7199) sequences within the SARS spike protein have been identified.

The coronavirus spike proteins (and other similar surface viral proteins) are thought to undergo a conformational change upon receptor binding to the target cell membrane. One or more of the hydrophobic Membrane Fusion peptides are thought to become exposed and inserted into the target membrane as a result of this conformational change. The free energy released

upon subsequent refolding of the spike protein to its most stable conformation is believed to play a role in the merger of the viral and cellular membranes.

One or more SARS HR sequences, preferably HR2, or a fragment thereof may be used to inhibit viral entry and membrane fusion with a target mammalian host cell. The invention provides a method of inhibiting viral infection comprising administering a composition comprising one or more SARS HR polypeptides or a fragment thereof. Preferably, the composition comprises a SARS HR2 sequence.

In another embodiment, the invention includes a composition comprising a SARS HR1 sequence, or a fragment thereof and a SARS HR2 sequence, or a fragment thereof. The HR1 and HR2 sequences may optionally be associated together in an oligomer. The composition may comprise the intermediate domain sequence between the HR1 and HR2 domains. The use of such an intermediate sequence may facilitate oligomerization or other structural interaction between the HR regions.

HR sequences for use in the invention may be produced recombinantly by methods known in the art. The SARS HR sequences may be modified to facilitate bacterial expression. In particular, the HR sequences may be modified to facilitate transport of the recombinant protein to the surface of the bacterial host cell. For example, leader sequences to a bacterial membrane protein may be added to the N terminus of the recombinant HR sequences. HR sequences for use in the invention may alternatively be produced by chemical synthesis by methods known in the art (see below).

As discussed in more detail later in the specification, Applicants have identified structural similarities between the SARS spike protein and the surface protein of *Neisseria meningitidis*, NadA (and other similar bacterial adhesion proteins). Another means of facilitating bacterial expression of HR sequences includes the addition of the stalk and/or anchor sequences of a NadA-like protein to the C-terminus of the recombinant HR sequences. Recombinant sequences containing the bacterial anchor sequence may preferably be prepared in outer membrane vesicles (the preparation of which is discussed in more detail later in the application). Recombinant sequences missing the bacterial anchor sequences may be secreted and isolated from the supernatant.

The invention includes a polypeptide sequence comprising a first sequence and a second sequence, wherein said first sequence comprises a leader sequence for a bacterial membrane protein and wherein said second sequence comprises a HR sequence of a coronavirus. Preferably, said first sequence comprises the leader sequence for a bacterial adhesin protein. More preferably, said bacterial adhesion protein is NadA. Preferably said second sequence comprises HR1, HR2 or both. In one embodiment, the second sequence comprises HR1, HR2 and the intermediate domain sequence present in the naturally occrding spike protein. For

example, the second sequence may comprise a fragment of a coronavirus spike protein comprising the amino acids starting with the N-terminus of the HR1 region and ending with the C-terminus of the HR2 region.

The invention further includes a polypeptide sequence comprising a first, second, third and fourth sequence, wherein the first sequence comprises a leader sequence for a bacterial membrane protein; wherein said second sequence comprises a HR sequence of a coronavirus; wherein said third sequence comprises a stalk domain of a bacterial adhesion protein; and wherein said fourth sequence comprises an anchor domain of a bacterial adhesion protein. In one embodiment, the first sequence comprising the leader peptide sequence is removed. In another embodiment, the third sequence comprising the stalk domain is removed. In another embodiment, the fourth sequence comprising the anchor domain is removed.

The polypeptide sequences of the above described constructs may be linked together by means known in the art, including, for example, via glycine linkers.

Examples of constructs which may be used in such bacterial expression systems are shown in FIGURE 50. Polypeptide sequences of each of the constructs illustrated in FIGURE 50 are given as SEQ ID NOS: 7200 to 7206.

7200 Leader NadA (1-29) - HR1 (879-980) - 6Xgly - HR2 (1144-1195) - stalk+anchor NadA (88-405)
7201 Leader NadA (1-29) - HR1 (879-980) - 6Xgly - HR2 (1144-1196) - stalk NadA (88-351)
7202 Leader NadA (1-29) - HR1 - HR2 (879-1196) - stalk+anchor NadA (88-405)
7203 Leader NadA (1-29) - HR1 - HR2 (879-1196)-stalk NadA (88-351)
7204 HR1 - HR2 (879-1196)-stalk NadA (88-351)-6xhis
7205 Leader NadA (1-29) - HR1 - HR2 (879-1196)-anchor NadA (351-405)
7206 Leader NadA (1-29) - HR1 - HR2 (879-1196)

Administration of one of more of these Membrane Fusion sequences may also interfere with the ability of a coronavirus to fuse to a host cell membrane. Accordingly, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 7197, SEQ ID NO: 7198and SEQ ID NO: 7199. The invention further includes an isolated polypeptide comprising an amino acid sequence having sequence homology to an amino acid sequence selected from the group consisting of SEQ ID NO: 7197, SEQ ID NO: 7198and SEQ ID NO: 7199.

Two or more of these SARS Membrane Fusion peptides can be combined together. The invention includes a composition comprising two SARS Membrane Fusion peptides wherein said peptides are selected from at least two of the amino acids selected from the group consisting of SEQ ID NO: 7197, SEQ ID NO: 7198and SEQ ID NO: 7199, or a sequence having sequence identity thereto.

Two or more of the SARS Membrane Fusion peptides may be linked together. Accordingly, the invention includes a polypeptide comprising a first amino acid sequence and a

second amino acid sequence, wherein said first and second amino acid sequences are selected from the group consisting of SEQ ID NO: 7197, SEQ ID NO: 7198and SEQ ID NO: 7199, or a sequence having sequence identity thereto. Preferably, said first amino acid sequence and said second amino acid sequence are different SARS Membrane Fusion peptides, *i.e.*, they are not the same.

The invention also includes a method of treating or preventing SARS virus infection comprising administering one or more of the SARS Membrane Fusion peptide compositions described above.

As discussed above, the spike protein is capable of forming trimers. The invention further includes a polypeptide comprising SEQ ID NO: 6042 in trimeric form. The invention includes a composition comprising at least polypeptides wherein each polypeptide comprises at least the alpha-helical coiled region of a SARS virus spike protein. Preferably, the spike protein comprises SEQ ID NO: 6042 or a fragment thereof.

The invention further includes a composition comprising a SARS virus spike protein or a fragment thereof wherein said protein is associated with a transmembrane and wherein said fragment comprises the alpha-helical region of the SARS virus spike protein. Preferably, the composition comprises at least three SARS virus spike proteins or a fragment thereof, wherein the fragment comprises the alpha-helical region of the SARS virus spike protein.

The invention further includes an antibody which specifically binds to a trimeric form of SARS virus spike proteins. Preferably, the spike protein comprises SEQ ID NO: 6042 or a fragment thereof. The invention includes an antibody which specifically binds to a trimeric form of SARS virus spike proteins wherein said proteins are associated with a transmembrane.

The invention further includes an antibody which specifically binds to a monomeric form of SARS virus spike protein or a fragment thereof. Preferably, the antibody specifically binds to a monomeric form of SEQ ID NO: 6042 or a fragment thereof.

The invention further includes a small molecule which interferes with or disrupts the coiling of a SARS viral spike protein trimer.

The invention further includes an attenuated SARS virus for use as a vaccine wherein said attenuated virus contains a polynucleotide insertion, deletion or substitution which does not disrupt the trimeric conformation of the SARS virus spike protein. The invention further includes an attenuated SARS virus for use as a vaccine wherein said attenuated virus contains a polynucleotide insertion, deletion or substitution which does not disrupt the alpha-helical formation of the SARS virus spike protein.

The spike protein may be recombinantly produced. In one embodiment, the spike protein is expressed in virus like particles so that the protein is attached to a cell membrane. Such attachment may facilitate presentation of immunogenic epitopes of the spike protein. Preferably,

the alpha-helical portion of the spike protein is associated with the cell membrane. Preferably, the spike proteins form a trimer within the transmembrane region of attachment.

Predicted N-glycosylation sites of SEQ ID NO: 6042 are identified below:

Position			Potential	Jury agreement	NGlyc result
29	NYTQ	SEQ ID NO: 7207	0.7751	(9/9)	+++
65	NVTG	SEQ ID NO: 7208	0.8090	(9/9)	+++
109	NKSQ	SEQ ID NO: 7209	0.6081	(7/9)	+
119	NSTN	SEQ ID NO: 7210	0.7039	(9/9)	++
158	NCTF	SEQ ID NO: 7211	0.5808	(7/9)	+
227	NITN	SEQ ID NO: 7212	0.7518	(9/9)	+++
269	NGTI	SEQ ID NO: 7213	0.6910	(9/9)	++
318	NITN	SEQ ID NO: 7214	0.6414	(9/9)	++
330	NATK	SEQ ID NO: 7215	0.6063	(8/9)	+
357	NSTF	SEQ ID NO: 7216	0.5746	(8/9)	+
589	NASS	SEQ ID NO: 7217	0.5778	(6/9)	+
602	NCTD	SEQ ID NO: 7218	0.6882	(9/9)	++
699	NFSI	SEQ ID NO: 7219	0.5357	(7/9)	+
783	NFSQ	SEQ ID NO: 7220	0.6348	(9/9)	++
1080	NGTS	SEQ ID NO: 7221	0.5806	(7/9)	+
1116	NNTV	SEQ ID NO: 7222	0.5106	(5/9)	+
1176	NESL	SEQ ID NO: 7223	0.6796	(9/9)	++

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6042 wherein said fragment comprises one or more of the glycosylation sites identified above (SEQ ID NOS: 7207-7223). The invention further includes a polynucleotide encoding one or more of the fragments identified above. This glycosylation site can be covalently attached to a saccharide. Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6042 wherein said fragment comprises one or more of the glycosylation sites identified above and wherein said polypeptide is glycosylated at one or more of the sites identified above.

Predicted O-glycosylation sites are identified below:

Residue	No.	Potential	Threshold	Assignment
Thr	698	0.8922	0.7696	T
Thr	706	0.9598	0.7870	T
Thr	922	0.9141	0.7338	T
Ser	36	0.8906	0.7264	S
Ser	703	0.8412	0.7676	S

The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6042 wherein said fragment comprises one or more of the O-glycosylation sites identified above. The invention further includes a polynucleotide encoding one or more of the fragments identified above. The invention further includes a polypeptide comprising a fragment of SEQ ID NO: 6042 wherein said fragment comprises one or more of the O-glycosylation sites identified above and further wherein the polypeptide is covalently bonded to a saccharide at one or more of the included glycosylation sites.

The invention further includes a polypeptide comprising a fragment of SEQ ID NO: 6042 wherein said fragment comprises one or more of the N-glycosylation sites identified above and

further wherein said fragment comprises one or more of the O-glycosylation sites identified above.

The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6042 wherein said fragment does not include one or more of the glycosylation sites identified above. The invention also includes a polynucleotide encoding such a polypeptide.

Predicted phosphorylation sites of SEQ ID NO: 6042 are Ser-346, Tyr-195, and Tyr-723. Accordingly, the invention comprises a polypeptide comprising a fragment of SEQ ID NO: 6042 wherein said fragment comprises at least ten amino acid residues and wherein said fragment comprises one or more of the amino acids selected from the group consisting of Ser-346, Tyr-195, and Tyr-723. In one embodiment, one or more of the amino acids selected from the group consisting of Ser-346, Tyr-195, and Tyr-723 are phosphorylated.

Expression and functional characterization of the spike glycoprotein has been described by Xiao *et al.* (2003) *Biochem Biophys Res Comm* 312:1159-1164.

T-epitopes for SEQ ID NO: 6042 are identified in Table 16. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified as SEQ ID NOS: 8041-8280; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 8041-8280, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus. The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide comprising SEQ ID NO: 6040. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 6040. The invention includes a fragment of a polypeptide comprising SEQ ID NO: 6040.

The invention includes a polynucleotide encoding SEQ ID NO: 6040 or a fragment thereof. The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 6040 or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding SEQ ID NO: 6040 or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising SEQ ID NO: 6040 or a fragment thereof.

SEQ ID NO: 6040 demonstrates functional homology with a membrane protein of coronaviruses. Predicted transmembrane helices of SEQ ID NO: 6040 are identified below:

Predicted Transmembrane Helices

```
The sequence positions in brackets denominate the core region.
Only scores above 500 are considered significant.

    Inside to outside helices :   3 found
          from         to    score center
      27 (  30)  48 (  45)   1138     38
     137 ( 139) 153 ( 153)    486    146

    Outside to inside helices :   3 found
          from         to    score center
      28 (  31)  45 (  45)    819     38
      71 (  73)  90 (  90)    210     81
     136 ( 142) 156 ( 156)    272    149
```

The amino acid region with the highest predicted transmembrane helical region is from amino acid position 27 to 48 of SEQ ID NO: 6040. Such transmembrane regions are often difficult to express recombinantly. Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6040 wherein said fragment does not include the amino acid sequence between positions 27 to 48. The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6040 wherein said fragment does not include the amino acid sequence between positions 28 to 45. The invention also includes a polynucleotide sequence encoding any of the above-identified polypeptides.

SEQ ID NO: 6040 is predicted to be a hypothetical protein of the SARS virus. A prediction of the protein localization of SEQ ID NO: 6040 is set forth below. SEQ ID NO: 6040 is predicted to be located in one of the following locations: mitochondrial matrix space, microbody (peroxisome), nucleus, and mitochondrial inner membrane. SEQ ID NO: 6040 is predicted to be associated with an organelle inside an infected cell.

Accordingly, SEQ ID NO: 6040 is a target for screening of chemical inhibitors to the SARS virus. The invention includes a polypeptide comprises SEQ ID NO: 6040 or a fragment thereof. The invention includes a polynucleotide encoding the polypeptide sequence of SEQ ID NO: 6040 or a fragment thereof. The invention includes a method of screening SEQ ID NO: 6040 for an inhibitor. The invention includes the recombinant expression of SEQ ID NO: 6040 in a host cell. The invention includes a small molecule which prevents the polypeptide of SEQ

ID NO: 6040 from associating with an organelle inside of an infected cell. The invention includes a fusion protein wherein said fusion protein comprises SEQ ID NO: 6040.

```
PSORT --- Prediction of Protein Localization Sites
                                                    version 6.4(WWW)
SEQ ID NO: 6040          163 Residues
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 0
McG: Examining signal sequence (McGeoch)
     Length of UR:    9
     Peak Value of UR:    1.75
     Net Charge of CR: 1
     Discriminant Score:      -2.56
GvH: Examining signal sequence (von Heijne)
     Signal Score (-3.5): 1.94
     Possible cleavage site: 53
>>> Seems to have no N-terminal signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 1
ALOM new cnt: 0 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
ALOM: finding transmembrane regions (Klein et al.)
     count: 0  value:    1.32  threshold: -2.0
     PERIPHERAL  Likelihood =  1.32
     modified ALOM score:  -1.16
Gavel: Examining the boundary of mitochondrial targeting seq.
      motif at: 156
     HRSVTI
Discrimination of mitochondrial target seq.:
     notclr ( 0.88)
Rule: mitochondrial protein
Rule: mitochondrial protein
Rule: mitochondrial protein
Rule: mitochondrial protein

*** Reasoning Step: 2

KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
  (Gavel position 156)  from: 27  to: 44  Score:  5.0
Mitochondrial matrix?  Score:  0.36
SKL motif (signal for peroxisomal protein):
     pos: 99(163), count: 1   SKL
     SKL score (peroxisome):  0.3
Amino Acid Composition Tendency for Peroxisome:  -4.28
Peroxisomal proteins?   Status: notclr
Amino Acid Composition tendency for lysosomal proteins
     score:  0.02  Status: notclr
Modified score for lysosome: 0.152
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
     Found: pos: 132 (5) KRKR
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
Checking the RNA binding motif (nucleus or cytoplasm)
nuc modified.   Score:  0.60
Nuclear Signal   Status: notclr ( 0.30)
```

```
Checking CaaX motif..
Checking N-myristoylation..
Checking CaaX motif..

----- Final Results -----
mitochondrial matrix space --- Certainty= 0.480(Affirmative) < succ>
microbody (peroxisome) --- Certainty= 0.300(Affirmative) < succ>
nucleus --- Certainty= 0.300(Affirmative) < succ>
mitochondrial inner membrane --- Certainty= 0.188(Affirmative) < succ>
```

Predicted N-glycosylation sites of SEQ ID NO: 6040 are identified below.

Position	Potential	Jury agreement	NGlyc result
2 NKTG (SEQ ID NO: 7255)	0.7804	(9/9)	+++
106 NLTL (SEQ ID NO: 7256)	0.6123	(7/9)	+

Accordingly, the invention comprises a fragment of SEQ ID NO: 6040 wherein said fragment is at least ten amino acids and wherein said fragment comprises one or more of the asparagines from the amino acid positions of SEQ ID NO: 6040 selected from the group consisting of 2 and 106. The invention includes a fragment of SEQ ID NO: 6040 wherein said fragment comprises one or more amino acid sequences selected from the group consisting of SEQ ID NO: 7255 and SEQ ID NO: 7256. Preferably, the fragment comprises the amino acid sequence NKTG (SEQ ID NO: 7255).

The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6040 wherein said fragment does not include one or more of the glycosylation sites identified above. The invention also includes a polynucleotide encoding such a polypeptide.

T-epitopes for SEQ ID NO: 6040 are identified in Table 14. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 7640-7800; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 7640-7800, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus.

The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide comprising SEQ ID NO: 6041. SEQ ID NO: 6041 demonstrates functional homology with a portion of an ORF 1ab polyprotein. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 6041. The invention includes a fragment of a polypeptide comprising SEQ ID NO: 6041. The invention includes a polynucleotide sequence encoding an amino acid sequence having sequence identity to SEQ ID NO: 6041. The invention includes a polynucleotide encoding a fragment of a polypeptide comprising SEQ ID NO: 6041.

The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 6041 or a fragment therof. The invention includes a diagnostic kit comprising a polynucleotide encoding SEQ ID NO: 6041 or a fragment thereof. The invention includes an immunogenic composition comprising a polypeptide comprising SEQ ID NO: 6041 or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising SEQ ID NO: 6041 or a fragment thereof.

The polyproteins of coronaviruses are associated with enzymatic activity. Accordingly, SEQ ID NO: 6041 is a target for screening of chemical inhibitors to the SARS virus. The invention includes a polypeptide comprising SEQ ID NO: 6041 or a fragment thereof. The invention includes a polynucleotide encoding the polypeptide sequence of SEQ ID NO: 6041 or a fragment thereof. The invention includes a method of screening SEQ ID NO: 6041 for an inhibitor. The invention includes the recombinant expression of SEQ ID NO: 6041 in a host cell. The invention includes a small molecule which prevents the polypeptide of SEQ ID NO: 6041 from performing enzymative activity. The invention includes a fusion protein wherein said fusion protein comprises SEQ ID NO: 6041.

Predicted transmembrane or hydrophobic regions of SEQ ID NO: 6041 are identified below. Although the polyprotein of coronaviruses is proteolytically cleaved into numerous smaller proteins, hydrophobic domains in the polyprotein are known to mediate the membrane association of the replication complex and to be able to dramatically alter the architecture of host cell membranes. Accordingly, the hydrophobic domains of the polyprotein are targets for genetic mutation to develop attenuated SARS virus vaccines. The hydrophobic domains are also

targets for small molecule inhibitors of the SARS virus. The hydrophobic domains may also be used to generate antibodies specific to those regions to treat or prevent SARS virus infection.

Possible transmembrane helices of SEQ ID NO: 6041

The sequence positions in brackets denominate the core region.
Only scores above 500 are considered significant.

Inside to outside helices : 18 found

from		to		score	center
234	(234)	254	(250)	1046	241
256	(256)	272	(270)	252	263
319	(319)	334	(334)	227	327
503	(505)	522	(519)	405	512
613	(615)	633	(629)	619	622
677	(679)	703	(696)	467	689
849	(851)	869	(865)	229	858
1080	(1080)	1097	(1094)	306	1087
1147	(1149)	1163	(1163)	354	1156
1557	(1557)	1581	(1577)	817	1567
1954	(1954)	1971	(1971)	832	1964
2369	(2372)	2395	(2387)	300	2379
2513	(2513)	2532	(2529)	690	2522

Outside to inside helices : 14 found

from		to		score	center
239	(239)	254	(254)	924	247
239	(248)	272	(263)	468	256
311	(314)	334	(328)	267	321
499	(503)	522	(519)	485	512
617	(617)	634	(631)	425	624
849	(853)	872	(872)	572	864
1147	(1147)	1162	(1162)	765	1155
1564	(1564)	1581	(1579)	883	1572
1951	(1951)	1968	(1966)	657	1958
2513	(2522)	2539	(2537)	711	2529

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6041, wherein said fragment comprises an amino acid sequence including one or more of the hydrophobic transmembrane sequences identified above. The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6041 wherein said fragment comprises one or more of the following polypeptide sequences of SEQ ID NO: 6041: 234-254, 613-633, 1557-1581, 1954-1971, 2513-2532, 239-254, 1564-1581, 1951-1968, 2513-2539. Preferably, the fragment comprises one or more of the following polypeptide sequences of SEQ ID NO: 6041: 234-254 and 239-254. The invention also includes polynucleotides encoding each of the polypeptide fragments identified above.

The invention includes an attenuated SARS virus wherein said attenuated SARS virus contains an addition, deletion or substitution in the polynucleotides encoding for one of the hydrophobic domains identified above. The invention also includes a method for creating an attenuated SARS virus comprising mutating a SARS virus by adding, deleting or substituting the

viral genome of the SARS virus to alter the coding of one or more of the hydrophobic domains of SEQ ID NO: 6041 identified above.

The invention includes an antibody which specifically identifies one or more of the hydrophobic regions of SEQ ID NO: 6041 identified above. The invention includes a small molecule which binds to, interferes with the hydrophobicity of or otherwise disrupts one or more of the hydrophobic regions of SEQ ID NO: 6041 identified above.

Predicted N-glycosylation sites of SEQ ID NO: 6041 are identified below:

Position		Potential	Jury agreement	NGlyc result
571	NLSH (SEQ ID NO: 7257)	0.6598	(8/9)	+
835	NTSR (SEQ ID NO: 7258)	0.5762	(7/9)	+
958	NVTD (SEQ ID NO: 7259)	0.7494	(9/9)	++
1113	NISD (SEQ ID NO: 7260)	0.7259	(8/9)	+
1205	NSTL (SEQ ID NO: 7261)	0.6296	(9/9)	++
1460	NVTG (SEQ ID NO: 7262)	0.6844	(9/9)	++
1685	NHSV (SEQ ID NO: 7263)	0.5181	(5/9)	+
2029	NKTT (SEQ ID NO: 7264)	0.5423	(5/9)	+

Accordingly, the invention comprises a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO: 6041, wherein said fragment comprises one or more of the N-glycosylation sites identified above. The invention comprises a polypeptide comprising a fragment of SEQ ID NO: 6041 wherein said fragment comprises one or more of sequences SEQ ID NOS: 7257-7264. Preferably, the fragment comprises one or more of the sequences SEQ ID NOS: 7257, 7259, 7260, 7261 and 7262. The invention further comprises a polynucleotide encoding one or more of the polypeptides identified above.

The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6041 wherein said fragment does not include one or more of the glycosylation sites identified above. The invention also includes a polynucleotide encoding such a polypeptide.

T-epitopes for SEQ ID NO: 6041 are identified in Table 15. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 7801-8040; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 7801-8040, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a

CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus.

The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide sequence SEQ ID NO: 6043 or a fragment thereof. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 6043. The invention includes a polynucleotide sequence encoding the amino acid sequence of SEQ ID NO: 6043 or a fragment thereof.

Predicted transmembrane regions of SEQ ID NO: 6043 are set forth below.

```
Inside to outside helices :   4 found
      from         to     score center
  41 (  41)  56 (  56)   1789     49
  76 (  79)  99 (  99)   2142     89
 105 ( 105) 125 ( 125)   1250    115

Outside to inside helices :   3 found
      from         to     score center
  41 (  41)  59 (  56)   2053     49
  76 (  82)  98 (  96)   1580     89
 103 ( 105) 125 ( 123)   1257    115
```

The amino acid region with the highest predicted transmembrane helical region is from amino acid position 76 to 99 of SEQ ID NO: 6043. Such transmembrane regions are often difficult to express recombinantly. Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6043 wherein said fragment does not include one or more of the hydrophobic amino acid sequences identified above. Preferably, the fragment does not include the amino acids between positions 27 to 48. The invention also includes a polynucleotide sequence encoding any of the above-identified polypeptides.

SEQ ID NO: 6043 is predicted to be a hypothetical protein of the SARS virus. A prediction of the protein localization of SEQ ID NO: 6043 is set forth below. SEQ ID NO: 6043 is predicted to be located in one of the following locations: mitochondrial inner membrane, plasma membrane, Golgi body, and mitochondrial intermembrane space. SEQ ID NO: 6043 may be associated with an organelle inside an infected cell.

Accordingly, SEQ ID NO: 6043 is a target for screening of chemical inhibitors to the SARS virus. The invention includes a polypeptide comprises SEQ ID NO: 6043 or a fragment thereof. The invention includes a polynucleotide encoding the polypeptide sequence of SEQ ID

NO: 6043 or a fragment thereof. The invention includes a method of screening SEQ ID NO: 6043 for an inhibitor. The invention includes the recombinant expression of SEQ ID NO: 6043 in a host cell. The invention includes a small molecule which prevents the polypeptide of SEQ ID NO: 6043 from associating with an organelle inside of an infected cell. The invention includes a fusion protein wherein said fusion protein comprises SEQ ID NO: 6043.

```
PSORT --- Prediction of Protein Localization Sites for SEQ ID NO: 6043
                                                    version 6.4(WWW)
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 3
     Position of the most N-terminal TMS: 40 at i=2
MTOP: membrane topology (Hartmann et al.)
     I(middle): 47   Charge diffirence(C-N):  3.5
McG: Examining signal sequence (McGeoch)
     Length  of  UR:   12
     Peak Value of UR:   1.41
     Net Charge of CR: 0
     Discriminant Score:     -4.67
GvH: Examining signal sequence (von Heijne)
     Signal Score (-3.5): 3.44
     Possible cleavage site: 15
>>> Seems to have no N-terminal signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 1
ALOM new cnt: 2 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
ALOM: finding transmembrane regions (Klein et al.)
     count: 2  value:  -6.90  threshold: -2.0
     INTEGRAL    Likelihood = -6.90   Transmembrane   83 -  99 (  78 - 101)
     INTEGRAL    Likelihood = -5.04   Transmembrane   40 -  56 (  37 -  60)
     PERIPHERAL  Likelihood = -0.32
     modified ALOM score:   1.48
>>> Likely a Type IIIb membrane protein (Nexo Ccyt)
Gavel: Examining the boundary of mitochondrial targeting seq.
      motif at: 128
     MRCWLC
Discrimination of mitochondrial target seq.:
     notclr ( 0.76)
Rule: mitochondrial protein
Rule: mitochondrial protein
Rule: mitochondrial protein
Rule: mitochondrial protein

*** Reasoning Step: 2

Type IIIa or IIIb is favored for ER memb. proteins
KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
  (Gavel position 128)  from: 39  to: 56  Score: 11.5
>>> Seems to have an intramitochondrial signal
Mitochondrial inner membrane?  Score:  0.59
Mitochondrial intermemb.space?  Score:  0.22
SKL motif (signal for peroxisomal protein):
     pos: 92(274), count: 1   SHL
```

```
     SKL score (peroxisome):  0.3
Amino Acid Composition Tendency for Peroxisome:   4.78
Peroxisomal proteins?   Status: positive
Amino Acid Composition tendency for lysosomal proteins
     score:  1.16  Status: notclr
Type III proteins may be localized at Golgi
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
Checking the RNA binding motif (nucleus or cytoplasm)
Nuclear Signal   Status: negative ( 0.00)
Check the Number of TMSs for typeIII (plasma memb.)
Checking N-myristoylation..

----- Final Results -----
mitochondrial inner membrane --- Certainty= 0.664(Affirmative) < succ>
plasma membrane --- Certainty= 0.600(Affirmative) < succ>
Golgi body --- Certainty= 0.400(Affirmative) < succ>
mitochondrial intermembrane space --- Certainty= 0.362(Affirmative) < succ>
```

Predicted N- and O- glycosylation sites of SEQ ID NO: 6043 are identified below.

Position	Potential	Jury agreement	NGlyc result
227 NATF (SEQ ID NO: 7265)	0.6328	(7/9)	+

Residue	No.	Potential	Threshold	Assignment
Thr	28	0.9095	0.6280	T
Thr	32	0.8740	0.6595	T
Thr	34	0.9058	0.6655	T
Thr	170	0.6816	0.6600	T
Thr	267	0.9240	0.5779	T
Thr	268	0.7313	0.5708	T
Thr	269	0.9859	0.5583	T
Thr	270	0.8023	0.5492	T
Ser	27	0.6930	0.6091	S
Ser	252	0.6457	0.5977	S

Accordingly, the invention comprises a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO: 6043, wherein said fragment comprises the N-glycosylation sites or O-glycosylation sites identified above. The invention comprises a polypeptide comprising a fragment of SEQ ID NO: 6043 wherein said fragment comprises one or more of the N-glycosylation sites or O-glycosylation sites identified above. The invention further comprises a polynucleotide encoding one or more of the polypeptides identified above.

The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6043 wherein said fragment does not include one or more of the glycosylation sites identified above. The invention also includes a polynucleotide encoding such a polypeptide.

T-epitopes for SEQ ID NO: 6043 are identified in Table 17. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 8281-8486; (b) an amino acid sequence having sequence identity to an amino acid

sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 8281-8486, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus. The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide comprising SEQ ID NO: 6044. The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6044 or a sequence having sequence identity to SEQ ID NO:206. The invention includes a polynucleotide encoding SEQ ID NO: 6044.

SEQ ID NO: 6044 is identified as a hypothetical protein. Predicted hydrophobic or transmembrane regions of SEQ ID NO: 6044 are identified below:

```
Inside to outside helices :   3 found
      from         to    score center
  1 (   1)  17 (  15)    891      8
 47 (  47)  66 (  63)    221     56

Outside to inside helices :   4 found
      from         to    score center
  1 (   4)  21 (  19)    599     11
```

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6044 wherein said fragment does not include one or more of the hydrophobic amino acid sequences identified above. Preferably, the fragment does not include the amino acids between positions 1 to 19. The invention also includes a polynucleotide sequence encoding any of the above-identified polypeptides.

SEQ ID NO: 6044 is predicted to be a hypothetical protein of SARS virus. A prediction of the protein localization of SEQ ID NO: 6044 is set forth below. SEQ ID NO: 6044 is predicted to be located in one of the following locations: nucleus, mitochondrial matrix, lysosome (lumen),

and microbody (peroxisome). SEQ ID NO: 6044 may be associated with an organelle inside an infected cell.

Accordingly, SEQ ID NO: 6044 is a target for screening of chemical inhibitors to the SARS virus. The invention includes a polypeptide comprises SEQ ID NO: 6044 or a fragment thereof. The invention includes a polynucleotide encoding the polypeptide sequence of SEQ ID NO: 6044 or a fragment thereof. The invention includes a method of screening SEQ ID NO: 6044 for an inhibitor. The invention includes the recombinant expression of SEQ ID NO: 6044 in a host cell. The invention includes a small molecule which prevents the polypeptide of SEQ ID NO: 6044 from associating with an organelle inside of an infected cell. The invention includes a fusion protein wherein said fusion protein comprises SEQ ID NO: 6044.

```
PSORT --- Prediction of Protein Localization Sites for SEQ ID NO: 6044
                                                    version 6.4(WWW)

154 Residues
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 0
McG: Examining signal sequence (McGeoch)
     Length of UR:   7
     Peak Value of UR:   1.06
     Net Charge of CR: 1
     Discriminant Score:     -7.97
GvH: Examining signal sequence (von Heijne)
     Signal Score (-3.5): -3.28
     Possible cleavage site: 34
>>> Seems to have no N-terminal signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 1
ALOM new cnt: 0 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
ALOM: finding transmembrane regions (Klein et al.)
     count: 0  value:   1.43  threshold: -2.0
     PERIPHERAL  Likelihood =  1.43
     modified ALOM score:  -1.19
Gavel: Examining the boundary of mitochondrial targeting seq.
      motif at: 151
     FRKKQV
Discrimination of mitochondrial target seq.:
     notclr (-0.46)

*** Reasoning Step: 2

KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
  (Gavel position 151)  from: 46  to: 50  Score:  5.0
Mitochondrial matrix?  Score:  0.36
SKL motif (signal for peroxisomal protein):
     pos: -1(154), count: 0
Amino Acid Composition Tendency for Peroxisome:   0.61
Peroxisomal proteins?   Status: notclr
     AAC score (peroxisome): 0.149
Amino Acid Composition tendency for lysosomal proteins
```

```
     score:  0.81  Status: notclr
Modified score for lysosome: 0.231
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
     Found: pos: 134 (3) KHKK
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
     Found:  pos: 136 (3) KK VSTNLCTHSF RKKQV
Final Robbins Score (nucleus):  0.60
Checking the RNA binding motif (nucleus or cytoplasm)
nuc modified.   Score:  0.90
Nuclear Signal   Status: positive ( 0.70)
Checking CaaX motif..
Checking N-myristoylation..
Checking CaaX motif..

----- Final Results -----
nucleus --- Certainty= 0.880(Affirmative) < succ>
mitochondrial matrix space --- Certainty= 0.360(Affirmative) < succ>
lysosome (lumen) --- Certainty= 0.231(Affirmative) < succ>
microbody (peroxisome) --- Certainty= 0.149(Affirmative) < succ>
```

One predicted O-glycosylation site of SEQ ID NO: 6044 is identified at residue 4:

```
Residue No.   Potential Threshold Assignment
Thr    4       0.6839   0.6484       T
```

Accordingly, the invention comprises a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO: 6044, wherein said fragment comprises the O-glycosylation site identified above. The invention further comprises a polynucleotide encoding one or more of the polypeptides identified above.

The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6044 wherein said fragment does not include one or more of the glycosylation sites identified above. The invention also includes a polynucleotide encoding such a polypeptide.

T-epitopes for SEQ ID NO: 6044 are identified in Table 18. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 8487-8665; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 8487-8665, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS

virus. The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide sequence comprising SEQ ID NO: 6045. The invention includes a polypeptide sequence comprising an amino acid sequence having sequence identity to SEQ ID NO: 6045. The invention includes a polypeptide sequence comprising a fragment of SEQ ID NO: 6045. The invention includes a polynucleotide sequence encoding any of these polypeptides.

SEQ ID NO: 6045 demonstrates functional homology with the envelope or small membrane protein of coronaviruses. The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 6045 or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide encoding SEQ ID NO: 6045 or a fragment thereof. The invention includes an immunogenic composition comprising SEQ ID NO: 6045 or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising SEQ ID NO: 6045 or a fragment thereof.

Predicted transmembrane regions of SEQ ID NO: 6045 are identified below:

```
Inside to outside helices :   1 found
     from        to    score center
  17 (  19)  33 (  33)   2881     26

Outside to inside helices :   1 found
     from        to    score center
  17 (  17)  34 (  34)   2981     27
```

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6045 wherein said fragment does not include one or more of the hydrophobic amino acid sequences identified above. Preferably, the fragment does not include the amino acids between positions 17 to 34. The invention also includes a polynucleotide sequence encoding any of the above-identified polypeptides. In one embodiment, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6045 wherein said fragment does not include amino acid residues 1-34 of SEQ ID NO: 6045.

Predicted protein Localization Site of SEQ ID NO: 6045 is below.

```
PSORT --- Prediction of Protein Localization Sites for SEQ ID NO: 6045
                                                    version 6.4(WWW)
Species classification: 4

*** Reasoning Step: 1
```

```
Preliminary Calculation of ALOM (threshold: 0.5)
     count: 2
     Position of the most N-terminal TMS: 17 at i=1
MTOP: membrane topology (Hartmann et al.)
     I(middle): 24   Charge diffirence(C-N):  2.0
McG: Examining signal sequence (McGeoch)
     Length  of  UR:   29
     Peak Value of UR:   3.40
     Net Charge of CR: -2
     Discriminant Score:     13.07
GvH: Examining signal sequence (von Heijne)
     Signal Score (-3.5): 4.37
     Possible cleavage site: 32
... positive value of mtop ...
>>> Seems to have an uncleavable N-term signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 1
ALOM new cnt: 1 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
ALOM: finding transmembrane regions (Klein et al.)
     count: 1  value: -15.12  threshold: -2.0
     INTEGRAL    Likelihood =-15.12   Transmembrane   17 -  33 (   8 -  44)
     PERIPHERAL  Likelihood =  0.47
     modified ALOM score:   3.12
>>> Seems to be a Type Ib (Nexo Ccyt) membrane protein
     The cytoplasmic tail is from 34 to 76 (44 Residues)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway
( 6) or uncleavable?
Gavel: Examining the boundary of mitochondrial targeting seq.
      motif at: 6
     Uncleavable? Ipos set to: 16
Discrimination of mitochondrial target seq.:
     notclr ( 0.19)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway

*** Reasoning Step: 2

> Relative position of the end of the tail: 44%
Memb.protein with uncleavable signl is often at ER
KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
  (Gavel position 16)  from: 70  to: 99  Score: 21.5
>>> Seems to have an intramitochondrial signal
SKL motif (signal for peroxisomal protein):
     pos: -1(76), count: 0
Amino Acid Composition Tendency for Peroxisome:  -4.11
Peroxisomal proteins?   Status: negative
Amino Acid Composition tendency for lysosomal proteins
     score:  0.68  Status: notclr
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
Checking the RNA binding motif (nucleus or cytoplasm)
Nuclear Signal   Status: negative ( 0.00)
Check cytoplasmic tail for typeIb (plasma memb.)
```

```
Checking the NPXY motif..
Checking the YXRF motif..
Checking N-myristoylation..

----- Final Results -----
plasma membrane --- Certainty= 0.730(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty= 0.640(Affirmative) < succ>
endoplasmic reticulum (lumen) --- Certainty= 0.100(Affirmative) < succ>
outside --- Certainty= 0.100(Affirmative) < succ>
```

Predicted N-glycosylation sites of SEQ ID NO: 6045 are identified at residues 48 and 66:

Position	Potential	Jury agreement	NGlyc result	
48 NVSL	0.6514	(9/9)	++	(SEQ ID NO: 7266)
66 NSSE	0.5880	(7/9)	+	(SEQ ID NO: 7267)

Accordingly, the invention comprises a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO: 6045, wherein said fragment comprises one or more of the N-glycosylation sites identified above. The invention comprises a polypeptide comprising a fragment of SEQ ID NO: 6045 wherein said fragment comprises one or more of the N-glycosylation sites identified above. The invention further comprises a polynucleotide encoding one or more of the polypeptides identified above.

The invention includes a polypeptide comprising a fragment of SEQ ID NO: 6045 wherein said fragment does not include one or more of the glycosylation sites identified above. The invention also includes a polynucleotide encoding such a polypeptide.

T-epitopes for SEQ ID NO: 6045 are identified in Table 19. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 8666-8820; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 8666-8820, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus. The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide sequence comprising SEQ ID NO: 6046. The invention includes polypeptide sequences comprising an amino acid sequence having sequence identity to SEQ ID NO: 6046. The invention includes a polypeptide sequence comprising a fragment of SEQ ID NO: 6046. The invention includes a polynucleotide encoding one of these polypeptides.

SEQ ID NO: 6046 has functional homology with a matrix protein of a coronavirus. The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 6046 or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide encoding SEQ ID NO: 6046 or a fragment thereof. The invention includes an immunogenic composition comprising SEQ ID NO: 6046 or a fragment thereof. The invention includes an antibody which recognizes a polypeptide comprising SEQ ID NO: 6046 or a fragment thereof.

Predicted transmembrane regions of SEQ ID NO: 6046 are identified below.

```
Inside to outside helices :   3 found
     from         to     score center
  21 (  21)  38 (  36)   2412     29
  51 (  53)  69 (  69)   2645     60
  74 (  82)  96 (  96)   2464     89

Outside to inside helices :   3 found
     from         to     score center
  18 (  21)  38 (  38)   2363     28
  52 (  52)  67 (  67)   2363     60
  76 (  76)  95 (  92)   2605     84
```

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6046 wherein said fragment does not include one or more of the hydrophobic amino acid sequences identified above. Preferably, the fragment does not include the amino acids between positions selected from the group consisting of 18 to 38, 52 to 67 and 76 to 95. The invention also includes a polynucleotide sequence encoding any of the above-identified polypeptides.

Predicted protein localization of SEQ ID NO: 6046 is set forth below.

```
PSORT --- Prediction of Protein Localization Sites
                                                    version 6.4(WWW)
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 3
     Position of the most N-terminal TMS: 21 at i=1
MTOP: membrane topology (Hartmann et al.)
     I(middle): 28   Charge diffirence(C-N):  6.0
```

```
McG: Examining signal sequence (McGeoch)
     Length of UR:   1
     Peak Value of UR:   3.16
     Net Charge of CR: -3
     Discriminant Score:      2.21
GvH: Examining signal sequence (von Heijne)
     Signal Score (-3.5): 4.29
     Possible cleavage site: 39
... positive value of mtop ...
>>> Seems to have an uncleavable N-term signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 1
Cleavable signal was detected in ALOM?: 0B
ALOM: finding transmembrane regions (Klein et al.)
     count: 3  value:  -7.64  threshold:  0.5
     INTEGRAL    Likelihood = -7.64   Transmembrane   21 -  37 (  18 -  39)
     INTEGRAL    Likelihood = -7.59   Transmembrane   50 -  66 (  43 -  72)
     INTEGRAL    Likelihood = -5.04   Transmembrane   79 -  95 (  72 -  99)
     PERIPHERAL  Likelihood =  2.38
     modified ALOM score:   2.13
>>> Likely a Type IIIb membrane protein (Nexo Ccyt)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway
( 2) or uncleavable?
Gavel: Examining the boundary of mitochondrial targeting seq.
      motif at: 2
     Uncleavable? Ipos set to: 12
Discrimination of mitochondrial target seq.:
     negative (-4.16)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway

*** Reasoning Step: 2

Type IIIa or IIIb is favored for ER memb. proteins
Memb.protein with uncleavable signl is often at ER
KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
SKL motif (signal for peroxisomal protein):
     pos: -1(221), count: 0
Amino Acid Composition Tendency for Peroxisome:   5.01
Peroxisomal proteins?   Status: notclr
Amino Acid Composition tendency for lysosomal proteins
     score:  2.30  Status: positive
Type III proteins may be localized at Golgi
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
Checking the RNA binding motif (nucleus or cytoplasm)
Nuclear Signal   Status: negative ( 0.00)
Check the Number of TMSs for typeIII (plasma memb.)
Checking N-myristoylation..

----- Final Results -----
endoplasmic reticulum (membrane) --- Certainty= 0.685(Affirmative) < succ>
plasma membrane --- Certainty= 0.640(Affirmative) < succ>
Golgi body --- Certainty= 0.460(Affirmative) < succ>
endoplasmic reticulum (lumen) --- Certainty= 0.100(Affirmative) < succ>
```

One predicted N-glycosylation sites of SEQ ID NO: 6046 is identified at residue 4:

Prediction of N-glycosylation sites

Position	Potential	Jury agreement	NGlyc result	
4 NGTI	0.8430	(9/9)	+++	(SEQ ID NO: 7268)

Accordingly, the invention comprises a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO: 6046, wherein said fragment comprises the N-glycosylation site identified above. The invention further comprises a polynucleotide encoding one or more of the polypeptides identified above.

The invention further comprises a polypeptide comprising a fragment of amino acid sequence SEQ ID NO: 6046, wherein said fragment does not include the N-glycosylation site identified above. The invention includes a polynucleotide encoding such a fragment.

A variant of SEQ ID NO: 6046 that is included within the invention is SEQ ID NO: 9963. Compared to SEQ ID NO: 6046, this sequence has Val at residue 72 instead of Ala.

T-epitopes for SEQ ID NO: 6046 are identified in Table 20. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 8821-9018; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 8821-9018, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus. The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide sequence comprising SEQ ID NO: 6047 or a fragment thereof or an amino acid sequence having sequence identity thereto. Predicted transmembrane regions of SEQ ID NO: 6047 are identified below.

```
Inside to outside helices :    2 found
      from         to     score center
   7 (  10)   29 (  27)     729     17
  21 (  24)   41 (  41)     640     34

Outside to inside helices :    2 found
      from         to     score center
   4 (   4)   22 (  19)     874     12
  22 (  24)   41 (  41)     499     31
```

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6047 wherein said fragment does not include one or more of the hydrophobic amino acid sequences identified above. Preferably, the fragment does not include the amino acids between positions selected from the group consisting of 4 to 22 and 22 to 41. The invention also includes a polynucleotide sequence encoding any of the above-identified polypeptides.

SEQ ID NO: 6047 is predicted to be a hypothetical protein of SARS virus. A prediction of the protein localization of SEQ ID NO: 6047 is set forth below. SEQ ID NO: 6047 is predicted to be located in one of the following locations: plasma membrane, endoplasmic reticulum, Golgi body, and microbody (peroxisome). SEQ ID NO: 6047 may be associated with an organelle inside an infected cell or with viral entry to a host cell.

Accordingly, SEQ ID NO: 6047 is a target for screening of chemical inhibitors to the SARS virus. The invention includes a polypeptide comprises SEQ ID NO: 6047 or a fragment thereof. The invention includes a polynucleotide encoding the polypeptide sequence of SEQ ID NO: 6047 or a fragment thereof. The invention includes a method of screening SEQ ID NO: 6047 for an inhibitor. The invention includes the recombinant expression of SEQ ID NO: 6047 in a host cell. The invention includes a small molecule which prevents the polypeptide of SEQ ID NO: 6047 from associating with an organelle inside of an infected cell or interacting with a host cell membrane. The invention includes a fusion protein wherein said fusion protein comprises SEQ ID NO: 6047. Predicted protein localization of SEQ ID NO: 6047 is set forth below.

```
PSORT --- Prediction of Protein Localization Sites
                                                  version 6.4(WWW)
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 1
     Position of the most N-terminal TMS: 2 at i=1
MTOP: membrane topology (Hartmann et al.)
     I(middle): 9   Charge diffirence(C-N):  0.5
McG: Examining signal sequence (McGeoch)
```

```
      Length of UR:   6
      Peak Value of UR:   3.08
      Net Charge of CR: 0
      Discriminant Score:      5.12
GvH: Examining signal sequence (von Heijne)
      Signal Score (-3.5): -4.45
      Possible cleavage site: 34
>>> Seems to have an uncleavable N-term signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 1
ALOM new cnt: 1 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
ALOM: finding transmembrane regions (Klein et al.)
      count: 1  value:  -2.44  threshold: -2.0
      INTEGRAL    Likelihood = -2.44   Transmembrane    2 -  18 (   1 -  20)
      PERIPHERAL  Likelihood =  1.22
      modified ALOM score:   0.59
>>> Seems to be a Type II (Ncyt Cexo) membrane protein
      The cytoplasmic tail is from 1 to 1 (1 Residues)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway
( 5) or uncleavable?
Gavel: Examining the boundary of mitochondrial targeting seq.
      motif at: 5
      Uncleavable? Ipos set to: 15
Discrimination of mitochondrial target seq.:
      notclr ( 1.48)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway

*** Reasoning Step: 2

Relative position of the cytoplasmic tail: 1%
      Larger value (>30%) is favared for ER memb. proteins
Memb.protein with uncleavable signl is often at ER
KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
  (Gavel position 15)  from: 64  to: 93  Score: 30.0
>>> Seems to have an intramitochondrial signal
SKL motif (signal for peroxisomal protein):
      pos: -1(63), count: 0
Amino Acid Composition Tendency for Peroxisome:   1.91
Peroxisomal proteins?   Status: notclr
      AAC score (peroxisome): 0.161
Amino Acid Composition tendency for lysosomal proteins
      score:  0.04  Status: notclr
Checking the consensus for Golgi
Checking the consensus for Golgi
Checking the cytoplasmic tail of type II (Golgi)
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
Checking the RNA binding motif (nucleus or cytoplasm)
Nuclear Signal   Status: negative ( 0.00)
Check mitochondrial signal for typeII (plasma memb.)
Type II is favored for plasma memb. proteins
Checking the NPXY motif..
Checking the YXRF motif..
```

```
Checking N-myristoylation..

----- Final Results -----
plasma membrane --- Certainty= 0.685(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty= 0.640(Affirmative) < succ>
Golgi body --- Certainty= 0.370(Affirmative) < succ>
microbody (peroxisome) --- Certainty= 0.161(Affirmative) < succ>
```

T-epitopes for SEQ ID NO: 6047 are identified in Table 21. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 9019-9131; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 9019-9131, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus. The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide comprising SEQ ID NO: 6048, a fragment thereof or an amino acid sequence having sequence identity thereto. Predicted transmembrane regions of SEQ ID NO: 6048 are identified below.

```
Inside to outside helices :   2 found
      from         to     score center
   3 (   3)  18 (  18)    1857     10
 100 ( 100) 117 ( 115)    2904    107

Outside to inside helices :   2 found
      from         to     score center
   1 (   1)  15 (  15)    1299      8
 100 ( 100) 117 ( 115)    3009    107
```

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6048 wherein said fragment does not include one or more of the hydrophobic amino acid

sequences identified above. Preferably, the fragment does not include the amino acids between positions selected from the group consisting of 1 to 15 and 100 to 117. The invention also includes a polynucleotide sequence encoding any of the above-identified polypeptides.

SEQ ID NO: 6048 is predicted to be a hypothetical protein of SARS virus. A prediction of the protein localization of SEQ ID NO: 6048 is set forth below. SEQ ID NO: 6048 is predicted to be located in one of the following locations: plasma membrane, lysosome (membrane), microbody (peroxisome), and endoplasmic reticulum (membrane). SEQ ID NO: 6048 may be associated with an organelle inside an infected cell or may interact with a host cell plasma membrane during viral entry to the host cell.

Accordingly, SEQ ID NO: 6048 is a target for screening of chemical inhibitors to the SARS virus. The invention includes a polypeptide comprises SEQ ID NO: 6048 or a fragment thereof. The invention includes a polynucleotide encoding the polypeptide sequence of SEQ ID NO: 6048 or a fragment thereof. The invention includes a method of screening SEQ ID NO: 6048 for an inhibitor. The invention includes the recombinant expression of SEQ ID NO: 6048 in a host cell. The invention includes a small molecule which prevents the polypeptide of SEQ ID NO: 6048 from associating with an organelle inside of an infected cell or prevents the polypeptide from associating with the cell membrane of a host cell. The invention includes a fusion protein wherein said fusion protein comprises SEQ ID NO: 6048. Predicted protein localization of SEQ ID NO: 6048 is set forth below.

```
PSORT --- Prediction of Protein Localization Sites
                                                    version 6.4(WWW)
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 2
     Position of the most N-terminal TMS: 3 at i=2
MTOP: membrane topology (Hartmann et al.)
     I(middle): 10   Charge diffirence(C-N): -2.5
McG: Examining signal sequence (McGeoch)
     Length of UR:   13
     Peak Value of UR:   3.38
     Net Charge of CR: 1
     Discriminant Score:     10.02
GvH: Examining signal sequence (von Heijne)
     Signal Score (-3.5): 2.56
     Possible cleavage site: 15
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 16
ALOM new cnt: 2 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 1B
ALOM: finding transmembrane regions (Klein et al.)
     count: 1  value: -14.75  threshold: -2.0
     INTEGRAL    Likelihood =-14.75   Transmembrane  101 - 117 (  95 - 120)
     PERIPHERAL  Likelihood =  6.63
     modified ALOM score:   3.05
```

```
>>> Seems to be a Type Ia membrane protein
     The cytoplasmic tail is from 118 to 122 (5 Residues)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway
(15) or uncleavable?
Gavel: Examining the boundary of mitochondrial targeting seq.
      motif at: 15
     Uncleavable? Ipos set to: 25
Discrimination of mitochondrial target seq.:
     notclr ( 0.73)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway

*** Reasoning Step: 2

KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
  (Gavel position 25)  from: 3  to: 12  Score:  8.5
SKL motif (signal for peroxisomal protein):
     pos: -1(122), count: 0
Amino Acid Composition Tendency for Peroxisome:   2.46
     AAC not from the N-term., score modified
Peroxisomal proteins?   Status: notclr
     AAC score (peroxisome): 0.115
Amino Acid Composition tendency for lysosomal proteins
     score: -0.40  Status: negative
GY motif in the tail of typeIa? (lysosomal)
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
Checking the RNA binding motif (nucleus or cytoplasm)
Nuclear Signal   Status: negative ( 0.00)
Type Ia is favored for plasma memb. proteins
Checking the NPXY motif..
Checking the YXRF motif..
Checking N-myristoylation..
Checking GPI anchor..
>>> Seems to be GPI-anchored  (0.85)

----- Final Results -----
plasma membrane --- Certainty= 0.919(Affirmative) < succ>
lysosome (membrane) --- Certainty= 0.200(Affirmative) < succ>
microbody (peroxisome) --- Certainty= 0.115(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty= 0.100(Affirmative) < succ>
```

T-epitopes for SEQ ID NO: 6048 are identified in Table 22. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 9132-9308; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further

comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 9132-9308, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus. The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide comprising SEQ ID NO: 6049, a fragment thereof or an amino acid sequence having sequence identity thereto. Predicted transmembrane or hydrophobic regions of SEQ ID NO: 6049 are identified below.

```
Inside to outside helices :   1 found
      from         to     score center
  13 (  13)   30 (  28)    3532     20

Outside to inside helices :   1 found
      from         to     score center
   9 (  11)   29 (  26)    3395     19
```

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6049 wherein said fragment does not include one or more of the hydrophobic amino acid sequences identified above. The invention also includes a polynucleotide sequence encoding any of the above-identified polypeptides.

SEQ ID NO: 6049 is predicted to be a hypothetical protein of SARS virus. A prediction of the protein localization of SEQ ID NO: 6049 is set forth below. SEQ ID NO: 6049 is predicted to be located in one of the following locations: outside, microbody (peroxisome), endoplasmic reticulum (membrane) and endoplasmic reticulum (lumen). The highest ranking indicates that SEQ ID NO: 6049 is located on the outside of a cell. Accordingly, SEQ ID NO: 6049 may be a surface exposed protein.

Accordingly, SEQ ID NO: 6049 may be used in an immunogenic composition to raise an immune response against the SARS virus. It also may be used to generate antibodies specific to the SARS virus. Such antibodies may be used in a method of treatment or prevention of a SARS virus infection. Such antibodies may further be used in a diagnostic test to identify the presence or absence of SARS virus in a biological sample.

The invention includes a polypeptide comprises SEQ ID NO: 6049 or a fragment thereof. The invention includes a polynucleotide encoding the polypeptide sequence of SEQ ID NO: 6049 or a fragment thereof. The invention includes a method of screening SEQ ID NO: 6049 for an inhibitor. The invention includes the recombinant expression of SEQ ID NO: 6049 in a host cell. The invention includes a fusion protein wherein said fusion protein comprises SEQ ID NO: 6049. Predicted protein localization of SEQ ID NO: 6049 is set forth below.

```
PSORT --- Prediction of Protein Localization Sites
                                                    version 6.4(WWW)
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 1
     Position of the most N-terminal TMS: 11 at i=1
MTOP: membrane topology (Hartmann et al.)
     I(middle): 18   Charge diffirence(C-N): -2.0
McG: Examining signal sequence (McGeoch)
     Length of UR:   24
     Peak Value of UR:   3.69
     Net Charge of CR: -2
     Discriminant Score:     13.56
GvH: Examining signal sequence (von Heijne)
     Signal Score (-3.5): 0.52
     Possible cleavage site: 25
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 26
ALOM new cnt: 1 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 1B
ALOM: finding transmembrane regions (Klein et al.)
     count: 0  value:  14.80  threshold: -2.0
     PERIPHERAL  Likelihood = 14.80
     modified ALOM score:  -3.86
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway
( 2) or uncleavable?
Gavel: Examining the boundary of mitochondrial targeting seq.
     motif at: 2
     Uncleavable? Ipos set to: 12
Discrimination of mitochondrial target seq.:
     notclr ( 1.42)
Rule: vesicular pathway
Rule: vesicular pathway
Rule: vesicular pathway

*** Reasoning Step: 2

KDEL   Count: 0
Number of Potential N-glycosylation Sites: 0
Out: score 0.800
Checking apolar signal for intramitochondrial sorting
  (Gavel position 12)  from: 44  to: 73  Score: 30.0
>>> Seems to have an intramitochondrial signal
SKL motif (signal for peroxisomal protein):
     pos: -1(44), count: 0
```

```
Amino Acid Composition Tendency for Peroxisome:   9.47
     AAC not from the N-term., score modified
Peroxisomal proteins?   Status: notclr
     AAC score (peroxisome): 0.320
Amino Acid Composition tendency for lysosomal proteins
     score: -6.47  Status: negative
Number of NX(S/T) motif: 0
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
Checking the RNA binding motif (nucleus or cytoplasm)
Nuclear Signal   Status: negative ( 0.00)
Checking CaaX motif..
Checking N-myristoylation..
Checking CaaX motif..

----- Final Results -----
outside --- Certainty= 0.820(Affirmative) < succ>
microbody (peroxisome) --- Certainty= 0.320(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty= 0.100(Affirmative) < succ>
endoplasmic reticulum (lumen) --- Certainty= 0.100(Affirmative) < succ>
```

T-epitopes for SEQ ID NO: 6049 are identified in Table 23. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 9309-9437; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 9309-9437, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus. The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide comprising SEQ ID NO: 6050 or a fragment thereof or an amino acid sequence having sequence identity thereto. Predicted transmembrane or hydrophobic regions are identified below.

```
Inside to outside helices :    1 found
      from         to     score center
  13 (  15)   32 (  30)     558     23

Outside to inside helices :    1 found
      from         to     score center
  16 (  16)   30 (  30)     364     23
```

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6050 wherein said fragment does not include one or more of the hydrophobic amino acid sequences identified above. The invention also includes a polynucleotide sequence encoding any of the above-identified polypeptides.

SEQ ID NO: 6050 is predicted to be a hypothetical protein of SARS virus. A prediction of the protein localization of SEQ ID NO: 6050 is set forth below. SEQ ID NO: 6050 is predicted to be located in one of the following locations: lysosome (lumen), mitochondrial matrix space, mitochondrial inner membrane, and mitochondrial intermembrane space. SEQ ID NO: 6050 may be associated with an organelle inside an infected cell during the viral replication cycle.

Accordingly, SEQ ID NO: 6050 is a target for screening of chemical inhibitors to the SARS virus. The invention includes a polypeptide comprises SEQ ID NO: 6050 or a fragment thereof. The invention includes a polynucleotide encoding the polypeptide sequence of SEQ ID NO: 6050 or a fragment thereof. The invention includes a method of screening SEQ ID NO: 6050 for an inhibitor. The invention includes the recombinant expression of SEQ ID NO: 6050 in a host cell. The invention includes a small molecule which prevents the polypeptide of SEQ ID NO: 6050 from associating with an organelle inside of an infected cell or prevents the polypeptide from associating with the cell membrane of a host cell. The invention includes a fusion protein wherein said fusion protein comprises SEQ ID NO: 6050. Predicted protein localization of SEQ ID NO: 6050 is set forth below.

```
PSORT --- Prediction of Protein Localization Sites
                                                    version 6.4(WWW)
MYSEQ           84 Residues
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 0
McG: Examining signal sequence (McGeoch)
     Length  of  UR:   3
     Peak Value of UR:   1.46
     Net Charge of CR: 2
     Discriminant Score:     -5.73
GvH: Examining signal sequence (von Heijne)
```

```
     Signal Score (-3.5): -0.12
     Possible cleavage site: 29
>>> Seems to have no N-terminal signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 1
ALOM new cnt: 0 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
ALOM: finding transmembrane regions (Klein et al.)
     count: 0  value:   8.43  threshold: -2.0
     PERIPHERAL  Likelihood =  8.43
     modified ALOM score:  -2.59
Gavel: Examining the boundary of mitochondrial targeting seq.
      motif at: 61
     ARCWYL
Discrimination of mitochondrial target seq.:
     positive ( 1.66)
Rule: mitochondrial protein
Rule: mitochondrial protein
Rule: mitochondrial protein
Rule: mitochondrial protein

*** Reasoning Step: 2

KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
  (Gavel position 61)  from: 52  to: 58  Score:  6.0
Mitochondrial matrix?  Score:  0.38
SKL motif (signal for peroxisomal protein):
     pos: -1(84), count: 0
Amino Acid Composition Tendency for Peroxisome:   1.47
Peroxisomal proteins?   Status: notclr
     AAC score (peroxisome): 0.263
Amino Acid Composition tendency for lysosomal proteins
     score:  2.86  Status: positive
Modified score for lysosome: 0.850
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
Checking the RNA binding motif (nucleus or cytoplasm)
Nuclear Signal   Status: negative ( 0.00)
Checking CaaX motif..
Checking N-myristoylation..
Checking CaaX motif..

----- Final Results -----
lysosome (lumen) --- Certainty= 0.850(Affirmative) < succ>
mitochondrial matrix space --- Certainty= 0.544(Affirmative) < succ>
mitochondrial inner membrane --- Certainty= 0.266(Affirmative) < succ>
mitochondrial intermembrane space --- Certainty= 0.266(Affirmative) < succ>
```

One predicted N-glycosylation sites of SEQ ID NO: 6050 is identified at residue 43:

Position	Potential	Jury agreement	NGlyc result	
43 NVTI	0.6713	(9/9)	++	(SEQ ID NO: 7269)

Accordingly, the invention comprises a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO: 6050 wherein said fragment comprises the N-glycosylation site

identified above. The invention further comprises a polynucleotide encoding one or more of the polypeptides identified above.

The invention further comprises a polypeptide comprising a fragment of amino acid sequence SEQ ID NO: 6050 wherein said fragment does not include the N-glycosylation site identified above. The invention includes a polynucleotide encoding such a fragment.

T-epitopes for SEQ ID NO: 6050 are identified in Table 24. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 9438-9538; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 9438-9538, or a polynucleotide encoding such a polypeptide.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus. The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a polypeptide sequence comprising SEQ ID NO: 6051 or a fragment thereof or an amino acid sequence having sequence identity thereto. The invention includes a polypeptide sequence comprising SEQ ID NO: 6052 or a fragment thereof or an amino acid sequence having sequence identity thereto.

SEQ ID NO: 6051 and SEQ ID NO: 6052 demonstrate functional homology with a nucleocapsid protein of a coronavirus. The invention includes a diagnostic kit comprising a polypeptide comprising SEQ ID NO: 6051, SEQ ID NO: 6052 or a fragment thereof. The invention includes a diagnostic kit comprising a polynucleotide encoding SEQ ID NO: 6051, SEQ ID NO: 6052 or a fragment thereof. The invention includes an immunogenic composition comprising SEQ ID NO: 6051, SEQ ID NO: 6052 or a fragment thereof. The invention includes

an antibody which recognizes a polypeptide comprising SEQ ID NO: 6051, SEQ ID NO: 6052 or a fragment thereof.

SEQ ID NO: 6051 is predicted to be phosphorylated at Ser-79; Thr-92; Ser-106; Thr-116; Thr-142; Ser-184; Ser-188; Ser-202; Ser-236; Thr-248; Ser-251; Ser-256; Thr-377. Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6051 wherein said fragment includes one or more of the amino acid residues of SEQ ID NO: 6051 selected from the group consisting of Ser-79; Thr-92; Ser-106; Thr-116; Thr-142; Ser-184; Ser-188; Ser-202; Ser-236; Thr-248; Ser-251; Ser-256; Thr-377. The invention further includes a polypeptide comprising a fragment of SEQ ID NO: 6051 wherein said fragment does not include one or more of the amino acid residues of SEQ ID NO: 6051 selected from the group consisting of Ser-79; Thr-92; Ser-106; Thr-116; Thr-142; Ser-184; Ser-188; Ser-202; Ser-236; Thr-248; Ser-251; Ser-256; Thr-377. Two further useful fragments of the N protein (*e.g.* for immunoassay) are SEQ ID NOS: 9783 & 9784, which are lysine-rich and can be used to distinguish the SARS virus from other coronaviruses.

Predicted transmembrane regions of SEQ ID NO: 6051 are identified below.

```
Inside to outside helices :    1 found
      from         to     score center
 304 ( 304) 323 ( 319)     495    312

Outside to inside helices :    2 found
      from         to     score center
304 ( 304) 319 ( 319)     597    312
```

Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 6051 wherein said fragment does not include one or more of the hydrophobic amino acid sequences identified above. The invention also includes a polynucleotide sequence encoding any of the above-identified polypeptides.

Predicted protein localization of SEQ ID NO: 6051 is set forth below. SEQ ID NO: 6051 is predicted to be localized near the nucleus, lysosome (lumen), mitochondrial matrix space, and microbody (peroxisome). The highest ranking is for localization near the nucleus. Coronavirus nucleocapsid proteins are known to bind to viral RNA. Coronavirus nucleocapsid proteins are also thought to be important for cell mediated immunity. Accordingly, the invention includes a polynucleotide comprising SEQ ID NO: 6051. The invention further includes a viral vector or particle suitable for in vivo delivery of the polynucleotide sequence comprising a SARS virus nuceocapsid polynucleotide sequence or a fragment thereof. In one embodiment, the polynucleotide comprises SEQ ID NO: 6051 or a fragment thereof. The invention further includes a method for eliciting a cell mediated immune response comprising delivering a polynucleotide encoding a SARS virus nucleocapsid protein or a fragment thereof to a mammal. In one embodiment, the polynucleotide comprising SEQ ID NO: 6051 or a fragment thereof.

The invention further includes a method of screening SEQ ID NO: 6051 for an inhibitor. The invention includes the recombinant expression of SEQ ID NO: 6051 in a host cell. The invention includes a small molecule which prevents the polypeptide of SEQ ID NO: 6051 from binding to SARS virus RNA during viral replication. The invention includes a fusion protein wherein said fusion protein comprises SEQ ID NO: 6051. Predicted protein localization of SEQ ID NO: 6051 is set forth below.

```
PSORT --- Prediction of Protein Localization Sites
                                                    version 6.4(WWW)
Species classification: 4

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
     count: 0
McG: Examining signal sequence (McGeoch)
     Length of UR:   3
     Peak Value of UR:   0.19
     Net Charge of CR: 0
     Discriminant Score:    -15.98
GvH: Examining signal sequence (von Heijne)
     Signal Score (-3.5): -6.36
     Possible cleavage site: 58
>>> Seems to have no N-terminal signal seq.
Amino Acid Composition of Predicted Mature Form:
   calculated from 1
ALOM new cnt: 0 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
ALOM: finding transmembrane regions (Klein et al.)
     count: 0  value:   5.04  threshold: -2.0
     PERIPHERAL  Likelihood =  5.04
     modified ALOM score:  -1.91
Gavel: Examining the boundary of mitochondrial targeting seq.
      motif at: 17
     PRITFG
Discrimination of mitochondrial target seq.:
     negative (-3.97)

*** Reasoning Step: 2

KDEL   Count: 0
Checking apolar signal for intramitochondrial sorting
Mitochondrial matrix?  Score:  0.10
SKL motif (signal for peroxisomal protein):
     pos: -1(399), count: 0
Amino Acid Composition Tendency for Peroxisome:   0.04
Peroxisomal proteins?   Status: notclr
     AAC score (peroxisome): 0.072
Amino Acid Composition tendency for lysosomal proteins
     score:  0.96  Status: notclr
Modified score for lysosome: 0.246
Checking the amount of Basic Residues (nucleus)
Checking the 4 residue pattern for Nuclear Targeting
     Found: pos: 256 (4) KKPR
     Found: pos: 372 (5) KKKK
Checking the 7 residue pattern for Nuclear Targeting
Checking the Robbins & Dingwall consensus (nucleus)
     Found:  pos: 372 (3) KK KKTDEAQPLP QRQKK
```

```
     Found:  pos: 373 (3) KK KTDEAQPLPQ RQKKQ
Final Robbins Score (nucleus):  0.80
Checking the RNA binding motif (nucleus or cytoplasm)
nuc modified.   Score:  0.90
Nuclear Signal   Status: positive ( 0.90)
Checking CaaX motif..
Checking N-myristoylation..
Checking CaaX motif..

----- Final Results -----
nucleus --- Certainty= 0.980(Affirmative) < succ>
lysosome (lumen) --- Certainty= 0.246(Affirmative) < succ>
mitochondrial matrix space --- Certainty= 0.100(Affirmative) < succ>
microbody (peroxisome) --- Certainty= 0.072(Affirmative) < succ>
```

Predicted N-glycosylation sites of SEQ ID NO: 6051 are identified below.

Position	Potential	Jury agreement	NGlyc result	
48 NNTA	0.6879	(9/9)	++	(SEQ ID NO: 7270)
270 NVTQ	0.7684	(9/9)	+++	(SEQ ID NO: 7271)

Residue No.	Potential	Threshold	Assignment
Thr 166	0.8547	0.6439	T
Thr 367	0.5575	0.5403	T
Thr 394	0.8217	0.5821	T

Accordingly, the invention comprises a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO: 6051 wherein said fragment comprises one or more of the N-glycosylation sites identified above. The invention further comprises a polynucleotide encoding one or more of the polypeptides identified above.

The invention further comprises a polypeptide comprising a fragment of amino acid sequence SEQ ID NO: 6051 wherein said fragment does not include one or more of the N-glycosylation sites identified above. The invention includes a polynucleotide encoding such a fragment.

T-epitopes for SEQ ID NO: 6052 are identified in Table 25. The invention includes a polypeptide for use as an antigen, wherein the polypeptide comprises: (a) an amino acid sequence selected from the group consisting of the T-epitope sequences identified in SEQ ID NOS: 9539-9752; (b) an amino acid sequence having sequence identity to an amino acid sequence of (a). The invention further comprising a polynucleotide sequence encoding the polypeptides of (a) or (b). The invention further comprising a method of expression or delivery of such polynucleotides through viral vectors and/or viral particles. The invention further comprises a polypeptide comprising two or more of the T-epitope sequences identified in SEQ ID NOS: 9539-9752, or a polynucleotide encoding such a polypeptide.

A variant of SEQ ID NO: 6052 that is included within the invention is SEQ ID NO: 9964. Compared to SEQ ID NO: 6052, this sequence has Ile at residue 54 instead of Thr.

The use as an antigen is preferably a use: (1) as a T-cell antigen; (2) for generating a complex between a class I MHC protein (*e.g.* a class I HLA) and a fragment of said antigen; (3) as an antigen for raising a cell-mediated immune response; and/or (4) as an antigen for raising a CTL response. The use preferably protects or treats disease and/or infection caused by a SARS virus. The invention provides the use of a polypeptide in the manufacture of a medicament for immunising a mammal (typically a human) against SARS viral infection wherein the polypeptide is as defined above.

The invention provides a method of raising an immune response in a mammal (typically a human), comprising the step of administering to the mammal a polypeptide as defined above, wherein said immune response is a cell-mediated immune response and, preferably, a CTL response. The immune response is preferably protective or therapeutic.

The invention includes a composition comprising a SARS virus nucleocapsid protein or a fragment thereof and further comprising a SARS virus membrane protein or a fragment thereof. The composition may further comprising one or more adjuvants discussed below.

The invention further includes a composition comprising a polypeptide comprising SEQ ID NO: 6051 or a fragment thereof or a sequence having sequence identity thereto and further comprising a polypeptide comprising SEQ ID NO: 6040, or a fragment thereof or a sequence having sequence identity thereto. Such composition may be used, for instance, in a vaccine. Such composition may further comprise one or more adjuvants discussed below.

The invention includes a composition comprising a SARS virus nucleocapsid protein or a fragment thereof and a SARS virus spike protein or a fragment thereof. In one embodiment the nucleocapsid protein comprises a polypeptide sequence comprising SEQ ID NO: 6051 or a fragment thereof or a sequence having sequence identity thereto. In one embodiment, the spike protein comprises a polynucleotide comprising SEQ ID NO: 6042 or a fragment thereof or a sequence having sequence identity thereto. The composition may further comprise one or more of the adjuvants discussed below.

The invention further includes a composition comprising antibodies specific to a SARS virus nucleocapsid protein and comprising antibodies specific to a SARS virus spike protein. In one embodiment the antibody is specific to a nucleocapsid protein comprises a polypeptide sequence comprising SEQ ID NO: 6051 or a fragment thereof or a sequence having sequence identity thereto. In one embodiment, the antibody is is specific to a spike protein comprises a polynucleotide comprising SEQ ID NO: 6042 or a fragment thereof or a sequence having sequence identity thereto.

The invention further includes polynucleotide sequences, and fragments thereof, of a SARS virus which are conserved among coronaviruses, and polypeptides encoded thereby. Such conserved sequences can be identified in the alignments shown in FIGURE 7. Such conserved

sequences may be used in the vaccines of the invention or in the diagnostic reagents, kits and methods of the invention.

The invention further includes polynucleotide sequences, and fragments thereof, of a SARS virus which are specific to SARS virus and not shared with coronaviruses. Such SARS specific sequences are also identified as SEQ ID NOS: 6040, 6043, 6044, 6047, 6048, 6049 and 6050. Such SARS specific sequences may be used in the vaccines of the invention or in the diagnostic reagents, kits and methods of the invention.

The invention also includes polynucleotide sequences which can be used as probes or primers for diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention includes a polynucleotide sequence comprising one or more of the primer sequences identified inSEQ ID NOS: 6076-6265 (Table 5). The invention further includes polynucleotide sequence comprising the complement of one or more of the primer sequences identified in SEQ ID NOS: 6076-6265.

The invention also includes polynucleotide sequences which can be used as probes or primers for diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention includes a polynucleotide sequence comprising one or more of the primer sequences identified in SEQ ID NOS: 6266-6343 (Table 6). The invention further includes polynucleotide sequence comprising the complement of one or more of the primer sequences identified in SEQ ID NOS: 6266-6343.

The invention also includes polynucleotide sequences which can be used as probes or primers for diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention includes a polynucleotide sequence comprising one or more of the primer sequences identified in SEQ ID NOS: 6344-6392 (Table 7). The invention further includes polynucleotide sequence comprising the complement of one or more of the primer sequences identified in SEQ ID NOS: 6344-6392..

The invention also includes polynucleotide sequences which can be used as probes or primers for diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention includes a polynucleotide sequence comprising one or more of the primer sequences identified in SEQ ID NOS: 6393-6559 (Tables 8 & 9). The invention further includes polynucleotide sequence comprising the complement of one or more of the primer sequences identified in SEQ ID NOS: 6393-6559.

The invention also includes polynucleotide sequences which can be used as probes or primers for diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention includes a polynucleotide sequence comprising one or more of the primer and probe sequences identified in SEQ ID NOS: 6560-6568. The invention further includes polynucleotide sequence comprising the complement of one or more of the primer sequences identified in SEQ ID NOS: 6560-6568.

The invention includes a polypeptide sequence comprising any one of even-numbered SEQ ID NOS: 7272-7290, or a fragment thereof, or a sequence having sequence identity thereto. The invention further includes a polynucleotide sequence encoding any one of even-numbered SEQ ID NOS: 7272-7290, or a fragment thereof, or a sequence having sequence identity thereto. Examples of such polynucleotide sequences are odd-numbered SEQ ID NOS: 7273-7291.

The invention includes a polynucleotide sequence comprising an intergenic sequence which is common to each open reading frame of the SARS virus. The SARS virus is thought to use this sequence to signal translation of the open reading frame. The intergenic sequence comprises a 10mer **SEQ ID NO: 7292**, or optionally a hexamer **SEQ ID NO: 7293**. When the virus transcribes its positive (+) RNA strand to (-) RNA strand, the virus replicating structure uses the (-) strand template to transcribe nucleotides at the 5' end prior to the first intergenic sequence, followed by the intergenic sequence, followed by the selected open reading frame. The virus then creates multiple mRNAs comprising the 5' end, the intergenic sequence and coding sequence. For more details on Nidovriales replication (including Coronavirus) see *e.g.*, Ziebuhr *et al.*, "Virus-encoded proteinases and proteolytic processing in the Nidovirales", *Journal of General Virology* 81:853-879 (2000), incorporated herein by reference in its entirety.

The invention comprising a polynucleotide sequence comprising SEQ ID NO: 7292 or the complement thereof. The invention comprising a polynucleotide sequence comprising SEQ ID NO: 7293 or the complement thereof. The invention further comprises a polynucleotide sequence comprising nucleotides from the 5' end of the SARS viral genome, or its reverse complement, and further comprising an intergenic sequence or its reverse complement. The polynucleotide may further comprise one or more of the SARS virus open reading frames. Examples of polynucleotide sequences comprising nucleotides from the 5' end of the SARS virus genome followed by the intergenic sequence are SEQ ID NOS: 7294-7301.-

The invention includes a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO: 7292, SEQ ID NO: 7293, SEQ ID NO: 7294, SEQ ID NO: 7295, SEQ ID NO: 7296, SEQ ID NO: 7297, SEQ ID NO: 7298, SEQ ID NO: 7299, SEQ ID NO: 7300 and SEQ ID NO: 7301, or a fragment thereof, or a sequence having sequence identity

thereto. In one embodiment, the polynucleotide does not consist entirely of a known SARS virus sequence.

The SARS virus intergenic sequence can be used to create a RNAi molecule. Such a SARS virus specific RNAi molecule can be used to treat SARS virus infection. The invention includes a RNAi molecule comprising a double stranded RNA molecule wherein one RNA strand comprises a sequence selected from the group consisting of SEQ ID NO: 7292, SEQ ID NO: 7293, SEQ ID NO: 7294, SEQ ID NO: 7295, SEQ ID NO: 7296, SEQ ID NO: 7297, SEQ ID NO: 7298, SEQ ID NO: 7299, SEQ ID NO: 7300 and SEQ ID NO: 7301, or a fragment thereof. Preferably, said RNA strand comprises a sequence selected from the group consisting of SEQ ID NO: 7292 and SEQ ID NO: 7293. Preferably, the other RNA strand comprises the reverse complement of the first strand or a polynucleotide sequence which hybridizes to the first strand.

The invention includes the use of RNAi in a method of treatment for SARS virus infection comprising administering to a mammal an effective amount of the si RNA molecule. Preferably, the RNAi molecule comprises the molecule described above. Further discussion of the RNAi applications of the intergenic sequence is included in section IV of the specification below.

The invention also includes the use of a SARS virus antisense nucleotide sequence, preferably antisense directed to the SARS virus intergenic sequence. Such an antisense sequence may be used in the treatment of a subject infected with the SARS virus. The antisense of the SARS virus intergenic sequence can be designed to bind to the SARS viral polynucleotides to block access of the viral replication machinery to the intergenic sequence. Such an antisense sequence may also be used to identify the presence or absence of a SARS virus in a biological sample. The antisence can itself be labeled or the antisense associated with viral polynucleotides can be detected by means known in the art.

Antisense nucleic acids are designed to specifically bind to RNA, resulting in the formation of RNA-DNA or RNA-RNA hybrids, with an arrest of DNA replication, reverse transcription or messenger RNA translation. Antisense polynucleotides based on a selected sequence can interfere with expression of the corresponding gene. Antisense polynucleotides will bind and/or interfere with the translation of the corresponding mRNA.

The invention also includes the use of the intergenic region with a ribozyme.

Trans-cleaving catalytic RNAs (ribozymes) are RNA molecules possessing endoribonuclease activity. Ribozymes are specifically designed for a particular target, and the target message must contain a specific nucleotide sequence. They are engineered to cleave any RNA species site-specifically in the background of cellular RNA. The cleavage event renders the mRNA unstable and prevents protein expression. Importantly, ribozymes can be used to inhibit

expression of a gene of unknown function for the purpose of determining its function in an in vitro or in vivo context, by detecting the phenotypic effect.

One commonly used ribozyme motif is the hammerhead, for which the substrate sequence requirements are minimal. Design of the hammerhead ribozyme is disclosed in Usman *et al.*, *Current Opin. Struct. Biol.* (1996) 6:527-533. Usman also discusses the therapeutic uses of ribozymes. Ribozymes can also be prepared and used as described in Long *et al.*, FASEB J. (1993) 7:25; Symons, *Ann. Rev. Biochem.* (1992) 61:641; Perrotta *et al.*, Biochem. (1992) 31:16-17; Ojwang *et al.*, *Proc. Natl. Acad. Sci.* (USA) (1992) 89:10802-10806; and US Patent 5,254,678. Ribozyme cleavage of HIV-I RNA is described in US Patent 5,144,019; methods of cleaving RNA using ribozymes is described in US Patent 5,116,742; and methods for increasing the specificity of ribozymes are described in US Patent 5,225,337 and Koizumi *et al.*, *Nucleic Acid Res.* (1989) 17:7059-7071. Preparation and use of ribozyme fragments in a hammerhead structure are also described by Koizumi *et al.*, *Nucleic Acids Res.* (1989) 17:7059-7071. Preparation and use of ribozyme fragments in a hairpin structure are described by Chowrira & Burke, *Nucleic Acids Res.* (1992) 20:2835. Ribozymes can also be made by rolling transcription as described in Daubendiek & Kool, *Nat. Biotechnol.* (1997) 15(3):273-277.

The hybridizing region of the ribozyme may be modified or may be prepared as a branched structure as described in Horn & Urdea, *Nucleic Acids Res.* (1989) 17:6959-67. The basic structure of the ribozymes may also be chemically altered in ways familiar to those skilled in the art, and chemically synthesized ribozymes can be administered as synthetic oligonucleotide derivatives modified by monomeric units. In a therapeutic context, liposome mediated delivery of ribozymes improves cellular uptake, as described in Birikh *et al.*, *Eur. J. Biochem.* (1997) 245:1-16.

Therapeutic and functional genomic applications of ribozymes proceed beginning with knowledge of a portion of the coding sequence of the gene to be inhibited. In the present invention, the target sequence preferably comprises the intergeneic sequence of the SARS virus. Preferably, the sequence is selected from the group consisting of SEQ ID NO: 7292 and SEQ ID NO: 7293. A target cleavage site is selected in the target sequence, and a ribozyme is constructed based on the 5' and 3' nucleotide sequences that flank the cleavage site. Preferably, the 5' nucleotide sequence includes the 5' untranslated region of the SARS virus. The ribozyme may then further be constructed from one or more of the polynucleotide sequences selected from the group consisting of SEQ ID NO: 7294, SEQ ID NO: 7295, SEQ ID NO: 7296, SEQ ID NO: 7297, SEQ ID NO: 7298, SEQ ID NO: 7299, SEQ ID NO: 7300 and SEQ ID NO: 7301.

Antisense treatment of HIV infection is described in the following references, each of which is incorporated herein by reference in their entirety. (antisense RNA complementary to the mRNA of gag, tat, rev, env) (Sezakiel *et al.*, 1991, J. Virol. 65:468-472; Chatterjee *et al.*,

1992, Science 258:1485-1488; Rhodes *et al.*, 1990, J. Gen. Virol. 71:1965. Rhodes *et al.*, 1991, AIDS 5:145-151; Sezakiel *et al.*, 1992, J. Virol. 66:5576-5581; Joshi *et al.*, 1991, J. Virol. 65:5524-5530).

The invention includes the use of decoy RNA to disrupt the SARS virus replication and life cycle. Methods of making and using such decoy RNA for treatment of a viral infection are known in the art. The invention includes delivery of genes encoding, for example, the SARS virus intergenic sequence, to infected cells. Preferably, the sequence comprises one or more of the sequences selected from the group consisting of SEQ ID NO: 7292, SEQ ID NO: 7293, SEQ ID NO: 7294, SEQ ID NO: 7295, SEQ ID NO: 7296, SEQ ID NO: 7297, SEQ ID NO: 7298, SEQ ID NO: 7299, SEQ ID NO: 7300 and SEQ ID NO: 7301. Preferably, the sequence comprises one or more of the sequences selected from the group consisting of SEQ ID NO: 7292 and SEQ ID NO: 7293. Preferably, the sequence comprises SEQ ID NO: 7293.

In the present invention, delivery of intergenic sequence which is not linked to the SARS virus open reading frames disrupts the translation process of the viral RNA and decreases the production of vial proteins. Similar methods of treatment for HIV viral infection have been described. The following references discuss the use of decoy RNA of HIV TAR or RRE for treatment of HIV infection. Each of these references is incorporated herein by reference in their entirety. (Sullenger *et al.*, 1990, Cell 63:601-608; Sullenger *et al.*, 1991, J. Virol. 65:6811-6816; Lisziewicz *et al.*, 1993, New Biol. 3:82-89; Lee *et al.*, 1994, J. Virol. 68:8254-8264), ribozymes (Sarver *et al.*, 1990, Science 247:1222-1225; Wecrasinghe *et al.*, 1991, J. Virol. 65:5531-5534; Dropulic *et al.*, 1992, J. Virol. 66:1432-1441; Ojwang *et al.*, 1992, Proc. Natl. Acad. Sci. USA. 89:10802-10806; Yu *et al.*, 1993, Proc. Natl. Acad. Sci. USA. 90:6340-6344; Yu *et al.*, 1995, Proc. Natl. Acad. Sci. USA. 92:699-703; Yamada *et al.*, 1994, Gene Therapy 1:38-45).

The invention includes the use of the SARS virus intergenic sequence in diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. Such diagnostic reagents, kits, and methods are further discussed in Section II of the specification.

The invention includes a pair of primers for amplifying a SARS polynucleotide sequence comprising (i) a first primer comprising a sequence which is substantially identical to a portion of a sequence selected from the group consisting of SEQ ID NO: 7292, SEQ ID NO: 7293, SEQ ID NO: 7294, SEQ ID NO: 7295, SEQ ID NO: 7296, SEQ ID NO: 7297, SEQ ID NO: 7298, SEQ ID NO: 7299, SEQ ID NO: 7300 and SEQ ID NO: 7301 and (ii) a second primer comprising a sequence which is substantially complementary to a portion of a sequence selected from the group consisting of the sequence SEQ ID NO: 1and the sequence SEQ ID NO: 2, such that the primer pair (i) and (ii) defines a template sequence within a sequence from the group consisting of the sequence SEQ ID NO: 1and the sequence SEQ ID NO: 2. Preferably, the (i)

first primer comprises a sequence which is substantially identical to a portion of a sequence selected from the group consisting of SEQ ID NO: 7292 and SEQ ID NO: 7293. Preferably, the (i) first primer comprises a sequence which is substantially identical to a portion of the sequence of SEQ ID NO: 7293. The amplicon defined by said first and second primers is preferably between 50 and 250 nucleotides in length. The primers may optionally be labeled to facilitate their detection. Methods and compositions for use in labeling primers are discussed further in the application in Section III.

The invention further includes a pair of primers for amplifying a SARS polynucleotide sequence comprising (i) a first primer comprising a sequence which is substantially identical to a portion of the complement of a portion of a sequence selected from the group consisting of SEQ ID NO: 7292, SEQ ID NO: 7293, SEQ ID NO: 7294, SEQ ID NO: 7295, SEQ ID NO: 7296, SEQ ID NO: 7297, SEQ ID NO: 7298, SEQ ID NO: 7299, SEQ ID NO: 7300 and SEQ ID NO: 7301 and (ii) a second primer comprising a sequence which is substantially complementary to a portion of the complement of a sequence selected from the group consisting of the sequence SEQ ID NO: 1and the sequence SEQ ID NO: 2, such that the primer pair defines a template sequence within a sequence selected from the group consisting of the sequence SEQ ID NO: 1 and the sequence SEQ ID NO: 2. The amplicon defined by said first and second primers is preferably between 50 and 250 nucleotides in length. The primers may optionally be labeled to facilitate their detection. Methods and compositions for use in labeling primers are discussed further in the application in Section III.

The invention includes a kit comprising (i) a first primer comprising a sequence which is substantially identical to a portion of a sequence selected from the group consisting of SEQ ID NO: 7292, SEQ ID NO: 7293, SEQ ID NO: 7294, SEQ ID NO: 7295, SEQ ID NO: 7296, SEQ ID NO: 7297, SEQ ID NO: 7298, SEQ ID NO: 7299, SEQ ID NO: 7300 and SEQ ID NO: 7301 and (ii) a second primer comprising a sequence which is substantially complementary to a portion of a sequence selected from the group consisting of the sequence SEQ ID NO: 1 and the sequence SEQ ID NO: 2, such that the primer pair (i) and (ii) defines a template sequence within a sequence from the group consisting of SEQ ID NO: 1 and SEQ ID NO: 2. Preferably, the (i) first primer comprises a sequence which is substantially identical to a portion of a sequence selected from the group consisting of SEQ ID NO: 7292 and SEQ ID NO: 7293. Preferably, the (i) first primer comprises a sequence which is substantially identical to a portion of the sequence of SEQ ID NO: 7293. The primers may optionally be labeled to facilitate their detection. Methods and compositions for use in labeling primers are discussed further in the application in Section III.

Other preferred kits comprise (i) a first primer comprising a sequence which is substantially identical to a portion of the complement of a portion of a sequence selected from

the group consisting of SEQ ID NO: 7292, SEQ ID NO: 7293, SEQ ID NO: 7294, SEQ ID NO: 7295, SEQ ID NO: 7296, SEQ ID NO: 7297, SEQ ID NO: 7298, SEQ ID NO: 7299, SEQ ID NO: 7300 and SEQ ID NO: 7301 and (ii) a second primer comprising a sequence which is substantially complementary to a portion of the complement of a sequence selected from the group consisting of SEQ ID NO: 1 and SEQ ID NO: 2, such that the primer pair defines a template sequence within a sequence selected from the group consisting of SEQ ID NO: 1 and SEQ ID NO: 2.

The invention further includes an attenuated SARS virus for use as a vaccine wherein the intergenic region has been mutated to reduce expression of the viral structural or nonstructural proteins. The attenuated SARS virus may comprises one or more additions, deletions or insertion in one or more of the intergenic regions of the viral genome. Preferably, the attenuated SARS virus comprises an addition, deletion or insertion in one or more occurrences of the sequence selected from the group consisting of SEQ ID NO: 7292 and SEQ ID NO: 7293. Preferably, the addition, deletion or insertion occurs in one or more occurrences of SEQ ID NO: 7293.

The invention further comprises a small molecule which inhibits binding or association of the SARS viral replication machinery, such as a ribonucleoprotein, with the intergenic region of the viral genome. Preferably, the small molecule inhibits binding or association of the SARS viral machinery with a sequence selected from the group consisting of SEQ ID NO: 7292 and SEQ ID NO: 7293. Preferably, the small molecule inhbiits binding or association of the SARS viral machinery with SEQ ID NO: 7293. The invention further includes a method of screening for a small molecule for treatment of SARS viral infection comprising using an assay to identify a small molecule which interferes with the association of the SARS viral replication machinery with the intergenic region of the SARS viral genome.

The invention further provides a novel SARS polynucleotide sequence SEQ ID NO: 9968. All six reading frames of this 690mer sequence are shown in Figure 113. The constituent amino acid sequences from Figure 113, having at least 4 amino acids, are listed as SEQ ID NOS: 9969 to 10032.

Accordingly the invention includes a polynucleotide sequence comprising SEQ ID NO: 9968. It also provides polynucleotide sequences having sequence identity to SEQ ID NO: 9968. The degree of sequence identity is preferably greater than 50% (*e.g.*, 60%, 70%, 80%, 85%, 88%, 90%, 92%, 95%, 99% or more).

The invention includes an amino acid sequence encoded by the polynucleotide sequence of SEQ ID NO: 9968, including the amino acid sequences selected from the group consisting of SEQ ID NO^{S}: 9969 to 10032. Preferably, the amino acid sequence comprises SEQ ID NO: 9997 or comprises SEQ ID NO: 9998.

The invention also provides amino acid sequences having sequence identity to an amino acid sequence encoded by SEQ ID NO: 9968. The invention provides amino acids having sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO^S: 9969 to 10032. The degree of sequence identity is preferably greater than 50% (e.g., 60%, 70%, 80%, 85%, 88%, 90%, 92%, 95%, 99% or more).

A portion of SEQ ID NO: 9968 matches with approximately 98% identity to a previously published SARS polynucleotide sequence, commonly referred to as "BNI-1" (SEQ ID NO: 10033). BNI-1 was sequenced at Bernhard Nocht Institute for Tropical Medicine, National Reference Center for Tropical Infectious Diseases in Hamburg, Germany. The BNI-1 sequence was published on the WHO website on April 4, 2003 at *http://www.who.int/csr/sars/primers/en* and in Dorsten *et al.*, "Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome", *New England Journal of Medicine*, published online at *http://www.nejm.org* on April 10, 2003. Both references are incorporated herein by reference in their entirety. The six reading frames of this 302mer sequence are shown in Figure 114 (see also Figure 129). The constituent amino acid sequences from Figure 114, having at least 4 amino acids, are listed as SEQ ID NO^S: 10034 to 10065. An alignment of SEQ ID NO: 10034 with SEQ ID NO: 9997 is shown in Figure 130.

The invention provides for polynucleotide sequences comprising fragments of SEQ ID NO: 9968. In one embodiment, the fragment does not consist entirely of SEQ ID NO: 10033 or of a known coronavirus.

The invention provides for amino acid sequences comprising fragments of an amino acid sequence encoded by SEQ ID NO: 9968. In one embodiment, the fragment does not consist entirely of an amino acid sequence encoded by SEQ ID NO: 10033 or a known coronavirus.

The invention provides for amino acids comprising fragments of an amino acid sequence selected from the group consisting of SEQ ID NO^S: 9969 to 10032. In one embodiment, the fragment does not consist entirely of an amino acid sequence encoded by the polynucleotide sequence of SEQ ID NO: 10033 or a known coronavirus.

Approximately 100 nucleotides at the 5' end of SEQ ID NO: 9968 do not match any portion of the BNI-1 polynucleotide sequence (SEQ ID NO: 10033). This unmatched portion is set forth as SEQ ID NO: 10066. The invention thus further provides a polynucleotide comprising the sequence comprising SEQ ID NO: 10066, polynucleotide sequences having sequence identity to SEQ ID NO: 10066, or polynucleotide sequences comprising fragments of SEQ ID NO: 10066.

The invention further comprises an amino acid sequence encoded by SEQ ID NO: 10066, an amino acid sequence having sequence identity to an amino acid sequence encoded by SEQ ID NO: 10066, or an amino acid sequence comprising fragments of an amino acid sequence

encoded by SEQ ID NO: 10066. Preferably, the amino acid sequence comprises SEQ ID NO: 10067.

SEQ ID NO: 9997/9998 demonstrates homology with the a region of pol1ab of several coronaviruses. FIGURE 115 shows an alignment of SEQ ID NO^S: 9997/9998 to amino acid sequences for pol1ab of bovine coronavirus (SEQ ID NO: 10068), avian infectious bronchitis virus (SEQ ID NO: 10069) and murine hepatitis virus (SEQ ID NO: 10070). A consensus amino acid sequence of SEQ ID NO^S: 9997/9998, SEQ ID NO: 10068, SEQ ID NO: 10069, and SEQ ID NO: 10070 is shown in the bottom row of the alignment in Figure 115 (*e.g.* SEQ ID NO: 10071).

As shown in FIGURE 113, the polynucleotide sequence encoding SEQ ID NO: 9997 has a stop codon after codon 205, between SEQ ID NO^S: 9997 and 9998. Optionally, the stop codon can be removed and the amino acid sequence continued (SEQ ID NO: 10072). Accordingly, the invention provides for an amino acid sequence comprising SEQ ID NO: 9997 and/or SEQ ID NO: 9998, or SEQ ID NO: 10072, and further comprising an amino acid sequence encoding for the C-terminus of a coronavirus pol1ab gene or a fragment thereof.

As shown in FIGURE 115, SEQ ID NO^S: 10068, 10069, 10070 and 10071 contain amino acids prior to the N-terminus of SEQ ID NO: 9997. The invention also provides for an amino acid sequence comprising SEQ ID NO: 9997 and further comprising an amino acid sequence encoding for the N-terminus of a coronavirus pol1ab protein or a fragment thereof.

The pol1ab sequences on FIGURE 115 contain a coding region indicated on the schematic of FIGURE 117 by a "*". In FIGURE 115, the beginning of this genomic region is designated by the arrow crossing in front of amino acid 6080 of the consensus sequence SEQ ID NO: 10071. The end of this genomic region is designated by the arrow crossing in front of amino acid 6604 of the consensus sequence. The invention provides for an amino acid sequence comprising SEQ ID NO: 9997 and/or SEQ ID NO: 9998, or SEQ ID NO: 10072, and further comprising a first amino sequence prior to the N-terminus of said SEQ ID NO: 9997 and/or SEQ ID NO: 9998, or SEQ ID NO: 10072, wherein said first amino acid sequence has homology to an N-terminus sequence of a known coronavirus pol1ab "*" protein or a fragment thereof.

The invention further provides for an amino acid sequence comprising SEQ ID NO: 9997 and SEQ ID NO: 9998, wherein the stop codon after SEQ ID NO: 9971 is removed (*i.e.* SEQ ID NO: 10072), and further comprising a second amino acid sequence following the C terminus of SEQ ID NO: 9998, wherein said second amino acid sequence is homologous with a C terminus of a known coronavirus pol1ab "*" protein or a fragment thereof.

Examples of such proteins are shown aligned in FIGURE 118, and are SEQ ID NO^S: 10073 to 10077. SEQ ID NO: 10073 comprises SEQ ID NO: 9997 and further comprises amino

acids prior to the N-terminus and subsequent to the C-terminus from the pol1ab "*" protein of avian infectious bronchitis virus. SEQ ID NO: 10074 comprises SEQ ID NO: 9997 and further comprises amino acids prior to the N-terminus and subsequent to the C-terminus from the pol1ab "*" protein of bovine coronavirus. SEQ ID NO: 10075 comprises SEQ ID NO: 9997 and further comprises amino acids prior to the N-terminus and subsequent to the C-terminus from the pol1ab "*" protein of murine hepatitis virus. SEQ ID NO: 10076 comprises SEQ ID NO: 9997 and further comprises amino acids prior to the N-terminus and subsequent to the C-terminus from the consensus of the pol1ab "*" protein of avian infectious bronchitis virus, bovine coronavirus, and murine hepatitis virus (FIGURE 115). SEQ ID NO: 10077 comprises the consensus sequence of SEQ ID NOS: 10073 to 10076.

The invention comprises an amino acid sequence selected from the group consisting of SEQ ID NO^S: 10073, 10074, 10075, 10076 and 10077. The invention further includes an amino acid sequence comprising fragments of an amino acid sequence selected from the group consisting of SEQ ID NO^S: 10073, 10074, 10075, 10076 and 10077. The invention further comprises an amino acid sequence with sequence identity to a sequence selected from the group consisting of SEQ ID NO^S: 10073, 10074, 10075, 10076 and 10077.

The invention comprises polynucleotides encoding for the amino acid sequences selected from the group consisting of SEQ ID NO^S: 10073, 10074, 10075, 10076 and 10077. The invention comprises polynucleotides having sequence identity to polynucleotides encoding for the amino acid sequences selected from the group consisting of SEQ ID NO^S: 10073, 10074, 10075, 10076 and 10077. The invention comprises fragments of polynucleotides encoding SEQ ID NO^S: 10073, 10074, 10075, 10076 and 10077.

As shown in Figure 113, SEQ ID NO: 9968 includes a sequence that encodes SEQ ID NO: 10020 followed by a stop codon, giving a C-terminus threonine (Thr) residue. The corresponding sequence from an amino acid sequence encoded by BNI-1 is SEQ ID NO: 10078, which continues past the C-terminus of SEQ ID NO: 10020. Accordingly, the invention includes a protein comprising amino acid sequence SEQ ID NO: 10020 or an amino acid sequence having sequence identity to SEQ ID NO: 10020 or an amino acid sequence comprising a fragment of SEQ ID NO: 10020, wherein the C-terminus residue of said protein is a threonine. Preferably, the C-terminus of said protein is –ST. Still more preferably, the C-terminus of said protein is –EST. The invention also includes a protein comprising amino acid sequence SEQ ID NO: 10078 or an amino acid sequence having sequence identity to SEQ ID NO: 10078 or an amino acid sequence comprising a fragment of SEQ ID NO: 10078, wherein the C-terminus residue of said protein is Thr. Preferably, the C-terminus of said protein is –ST. Still more preferably, the C-terminus of said protein is –EST.

SEQ ID NO: 9968 also encodes a 54mer amino acid sequence SEQ ID NO: 10015. The polynucleotide encoding SEQ ID NO: 10015 encodes two stop codons at its C-terminus (Figure 113). The corresponding region from the BNI-1 sequence does not contain this 54mer. Accordingly, the invention includes a protein comprising amino acid sequence SEQ ID NO: 10015, or an amino acid sequence having sequence identity to SEQ ID NO: 10015 or an amino acid sequence comprising a fragment of SEQ ID NO: 10015. The invention further includes a polypeptide comprising SEQ ID NO: 10015 and further comprising a first amino acid sequence prior to the N-terminus of SEQ ID NO: 10015.

SEQ ID NO: 9968 encodes the amino acid sequence SEQ ID NO: 9969. The polynucleotide sequence contains a stop codon at the C-terminus of SEQ ID NO: 9969. Accordingly, the invention includes a protein comprising amino acid sequence SEQ ID NO: 9969, or an amino acid sequence having sequence identity to SEQ ID NO: 9969. The invention further includes a polypeptide comprising SEQ ID NO: 9969 and further comprising a first amino acid sequence prior to the N-terminus of SEQ ID NO: 9969. The invention further includes a polypeptide comprising the sequence SEQ ID NO: 10079.

SEQ ID NO: 9968 encodes amino acid sequence QRT (Figure 113), followed by a stop codon. Accordingly, the invention includes a protein comprising amino acid sequence QRT. The invention further includes a polypeptide comprising amino acid sequence QRT and further comprising a first amino acid sequence prior to the N-terminus of the sequence QRT.

SEQ ID NO: 9968 encodes amino acid sequence SEQ ID NO: 10022, followed by a stop codon at its C-terminus. Accordingly, the invention includes a protein comprising amino acid sequence SEQ ID NO: 10022, or an amino acid sequence having sequence identity to SEQ ID NO: 10022. The invention further includes a polypeptide comprising SEQ ID NO: 10022 and further comprising a first amino acid sequence prior to the N-terminus of SEQ ID NO: 10022.

SEQ ID NO: 9968 encodes amino acid sequence SEQ ID NO: 10027. Within the SEQ ID NO: 10027 coding sequence there are at least three start codons, identified with underlining in Figure 119. The open reading frame indicated by the first start codon is SEQ ID NO: 10081. The open reading frame indicated by the second start codon is SEQ ID NO: 10082. The open reading frame indicated by the third start codon is SEQ ID NO: 10083.

The invention provides a novel SARS polynucleotide sequence SEQ ID NO: 10084. All six reading frames of this 1463mer sequence are shown in Figure 120 (see also Figure 122). The constituent amino acid sequences from Figure 120, having at least 4 amino acids, are listed as SEQ ID NOS: 10085 to 10209 (see Figures 120A to 120F).

The invention includes a polynucleotide sequence comprising SEQ ID NO: 10084. The invention also provides polynucleotide sequences having sequence identity to SEQ ID NO: 10084. The invention also provides for polynucleotide sequences comprising fragments of SEQ

ID NO: 10084. In one embodiment, the polynucleotide fragment does not consist entirely of SEQ ID NO: 10033 or a known coronavirus polynucleotide sequence or a known SARS polynucleotide sequence.

The invention includes an amino acid sequence encoded by the polynucleotide sequence of SEQ ID NO: 10084, including the amino acid sequences of Figures 120A to 120F *e.g.* selected from the group consisting of SEQ ID NO^S: 10085 to 10209. Preferably, the amino acid sequence comprises SEQ ID NO: 10149.

The invention also provides amino acid sequences having sequence identity to an amino acid sequence encoded by SEQ ID NO: 10084. The invention provides amino acids having sequence identity to an amino acid sequence from Figures 120A to 120F *e.g.* selected from the group consisting of SEQ ID NO^S: 10085 to 10209.

The invention also provides fragments of amino acid sequences encoded by SEQ ID NO: 10084. The invention also provides fragments of amino acid sequences selected from the group consisting of SEQ ID NO^S: 10085 to 10209. In one embodiment, the fragment does not consist entirely of an amino acid sequence encoded by SEQ ID NO: 10033 or an amino acid sequence of a known coronavirus or an amino acid sequence of a known SARS virus. An alignment of the matching portion of SEQ ID NO: 10033 and SEQ ID NO: 10084 is included in FIGURE 121.

In one embodiment, the invention comprises an amino acid sequence comprising SEQ ID NO: 10149. An alignment of the polynucleotide sequence SEQ ID NO: 10084 to the encoded SEQ ID NO: 10149 is shown in FIGURE 122 (5'3' Frame 3). Analysis of the 5'3' Frame 3 translation by a computer program to predict start codon methionines (NetStart 1.0) (FIGURE 123) reveals SEQ ID NO^S: 10210 to 10215.

The invention includes a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 10210, SEQ ID NO: 10211, SEQ ID NO: 10212, SEQ ID NO: 10213, SEQ ID NO: 10214 and SEQ ID NO: 10215. The invention includes a protein having sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO: 10210, SEQ ID NO: 10211, SEQ ID NO: 10212, SEQ ID NO: 10213, SEQ ID NO: 10214 and SEQ ID NO: 10215. In one embodiment, the protein does not consist entirely of an amino acid sequence of a known SARS virus or of a known coronavirus.

The invention includes a fragment of a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 10210, SEQ ID NO: 10211, SEQ ID NO: 10212, SEQ ID NO: 10213, SEQ ID NO: 10214 and SEQ ID NO: 10215. In one embodiment, the fragment does not consist entirely of an amino acid sequence of a known SARS virus or of a known coronavirus.

In one embodiment, the invention includes a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 10210, SEQ ID NO: 10211 and

SEQ ID NO: 10212. Partial results of a BLAST of SEQ ID NO: 10210 against GenBank is included in FIGURE 124. These results indicate that SEQ ID NOS: 10210, 10211 and 10212 have functional similarities to a Coronavirus RNA polymerase, particularly the RNA polymerase of murine hepatitis virus, bovine coronavirus, and avian infectious bronchitis.

In one embodiment, the invention is directed to a polypeptide comprising a first amino acid sequence selected from the group consisting of SEQ ID NO: 10210, SEQ ID NO: 10211 and SEQ ID NO: 10212 and a second amino acid sequence from the C-terminus of a coronavirus ORF1ab sequence. Preferably, the second amino acid sequence is from a bovine coronavirus. One example of this embodiment is shown below as SEQ ID NO: 10216. Amino acids 1-481 of SEQ ID NO: 10216 are the first amino acid sequence of SEQ ID NO: 10210, and amino acids 482-1152 are the second amino acid sequence of the C-terminus of a bovine coronavirus orf1ab polyprotein (Gi 26008080) (NP_150073.2) (SEQ ID NO: 10217).

Accordingly, the invention includes a polypeptide comprising SEQ ID NO: 10216. The invention further includes a polypeptide comprising a first amino acid sequence of SEQ ID NO: 10210 and a second amino acid sequence of SEQ ID NO: 10217. The invention further includes a polypeptide comprising a first amino acid sequence having greater than *x*% identity to SEQ ID NO: 10210 and a second amino acid sequence having greater than *y*% identity to SEQ ID NO: 10217, wherein *x* is greater than or equal to 85% (*e.g.*, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) and wherein *y* is greater than or equal to 60% (*e.g.*, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more).

The invention also includes a polypeptide comprising a fragment of SEQ ID NO: 10210, wherein said fragment includes an epitope. Computer-predicted epitopes of SEQ ID NO: 10210, using a 17mer window, are included in FIGURE 125A (Hopp & Woods) and FIGURE 125B (Kyte & Doolittle).

The amino acid sequence of SEQ ID NO: 10210 also contains two predicted glycosylation sites at amino acids 81–84 (NNTE; SEQ ID NO: 10218) and at 180–183 (NHSV; SEQ ID NO: 10219). Accordingly, the invention includes a polypeptide comprising a fragment of SEQ ID NO: 10210, wherein said fragment includes a glycosylation site. The invention further includes a polypeptide comprising a fragment of SEQ ID NO: 10210, wherein said fragment includes the Asn at position 81. Preferably, said Asn is glycosylated. The invention further includes a polypeptide comprising a fragment of SEQ ID NO: 10210, wherein said fragment includes the Asn at position 180. Preferably, said Asn is glycosylated.

In one embodiment, the invention includes a polypeptide comprising an amino acid sequence from within Figure 120D and/or SEQ ID NO^S: 10150 to 10160 *e.g.* from SEQ ID NO^S: 10154, 10155, 10158 and 10160. Within SEQ ID NO: 10154 the following amino acid sequences starting with a Met and ending at a stop codon can be identified: SEQ ID NO^S: 10220 to 10227.

Accordingly, the invention includes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 10220, SEQ ID NO: 10221, SEQ ID NO: 10222, SEQ ID NO: 10223, SEQ ID NO: 10224, SEQ ID NO: 10225, SEQ ID NO: 10226 and SEQ ID NO: 10227, or a fragment thereof or an amino acid sequence having sequence identity thereto.

In one embodiment, the invention includes a polypeptide comprising the amino acid sequence within Figure 120E *e.g.* from SEQ ID NO^S: 10161 to 10182, and in particular SEQ ID NOS: 10171 and 10176. Within SEQ ID NO^S: 10171 and 10176 the following amino acid sequences starting with a Met and ending at a stop codon can be identified: SEQ ID NO: 10228 and SEQ ID NO: 10229.

Accordingly, the invention includes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 10228 and SEQ ID NO: 10229, or a fragment thereof or an amino acid sequence having sequence identity thereto.

In one embodiment, the invention includes a polypeptide comprising an amino acid sequence from Figure 120F *e.g.* SEQ ID NO^S: 10183 to 10209. Within Figure 120F the following amino acid sequence starting with a Met and ending at a stop codon can be identified: SEQ ID NO: 10187. Accordingly, the invention includes a polypeptide comprising an amino acid sequence of SEQ ID NO: 10187, or a fragment thereof or an amino acid sequence having sequence identity thereto.

In one embodiment, the polynucleotides of the invention do not include one of the following primers, disclosed at *http://content.nejm.org/cgi/reprint/NEJMoa030781v2.pdf*:

5'GGGTTGGGACTATCCTAAGTGTGA3' (SEQ ID NO: 10230)
5'TAACACACAACICCATCATCA3' (SEQ ID NO: 10231)
5'CTAACATGCTTAGGATAATGG3' (SEQ ID NO: 10232)
5'GCCTCTCTTGTTCTTGCTCGC3' (SEQ ID NO: 10233)
5'CAGGTAAGCGTAAAACTCATC3' (SEQ ID NO: 10234)

The invention also includes polynucleotide sequences which can be used as probes for diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention includes the polynucleotide primers identified in Table 31 (SEQ ID NO^S: 10235 to 10258), the forward primers SEQ ID NO^S: 10259 to 10281 and the reverse primers SEQ ID NO^S: 10282 to 10298. The invention further includes polynucleotide sequences which are complementary to any one of these primer sequences disclosed herein.

The invention provides a SARS polynucleotide sequence SEQ ID NO: 10299. All six reading frames of this sequence are included in FIGURE 126 (See also Figure 131). The constituent amino acid sequences from Figure 126, having at least 4 amino acids, are listed as SEQ ID NOS: 10300 to 10337.

Accordingly, the invention includes a polynucleotide sequence comprising SEQ ID NO: 10299. It also provides polynucleotide sequences having sequence identity to SEQ ID NO: 10299. The invention also provides for polynucleotide sequences comprising fragments of SEQ ID NO: 10299. In one embodiment, the polynucleotide fragment does not consist entirely of a known polynucleotide sequence of a SARS virus or a known polynucleotide sequence of a coronavirus.

The invention includes an amino acid sequence encoded by the polynucleotide sequence of SEQ ID NO: 10299, including the amino acid sequences shown in Figure 126, and the amino acid sequences selected from the group consisting of SEQ ID NO^S: 10300 to 10337. Preferably, the amino acid sequence comprises SEQ ID NO: 10316.

The invention also provides amino acid sequences having sequence identity to an amino acid sequence encoded by SEQ ID NO: 10299. The invention provides amino acid sequences having identity to an amino acid sequence selected from the group consisting of SEQ ID NO^S: 10300 to 10337.

The invention also provides fragments of amino acid sequences encoded by SEQ ID NO: 10299. The invention also provides fragments of amino acid sequences selected from the group consisting of SEQ ID NO^S: 10300 to 10337. In one embodiment, the fragment does not consist entirely of a known amino acid sequence of a SARS virus or a known amino acid sequence of a coronavirus.

In one embodiment, the invention comprises an amino acid sequence comprising SEQ ID NO: 10316. Encoded open reading frames within SEQ ID NO: 10316 include SEQ ID NO: 10338 and SEQ ID NO: 10339.

In one embodiment, the invention comprises an amino acid sequence comprising a sequence from within the 5'3' Frame 1 translation of SEQ ID NO: 10299. The following encoded open reading frame is found within this translation: SEQ ID NO: 10340.

In one embodiment, the invention comprises an amino acid sequence comprising a sequence from within the 3'5' Frame 1 translation of SEQ ID NO: 10299. An encoded open reading frame within this translation is SEQ ID NO: 10341.

In one embodiment, the invention comprises an amino acid sequence comprising a sequence from within the 3'5' Frame 2 translation of SEQ ID NO: 10299. An encoded open reading frame within this translation is SEQ ID NO: 10342.

The invention includes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 10338, SEQ ID NO: 10339, SEQ ID NO: 10340, SEQ ID NO: 10341 and SEQ ID NO: 10342. The invention includes a polypeptide having sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO: 10338, SEQ ID NO: 10339, SEQ ID NO: 10340, SEQ ID NO: 10341 and SEQ ID NO: 10342. The invention

includes a fragment of a polypeptide comprising an amino acid sequence elected from the group consisting of SEQ ID NO: 10338, SEQ ID NO: 10339, SEQ ID NO: 10340, SEQ ID NO: 10341 and SEQ ID NO: 10342. In one embodiment, the fragment does not consist entirely of a known SARS virus amino acid sequence or of a known coronavirus amino acid sequence.

In one embodiment, SEQ ID NOS: 10338-10342 are used in fusion proteins. Accordingly, the start codon methionines may be removed. The invention comprises a amino acid sequence selected from the group consisting of SEQ ID NO: 10343, SEQ ID NO: 10344, SEQ ID NO: 10345, SEQ ID NO: 10346 and SEQ ID NO: 10347.

In one embodiment, the invention comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 10338 and SEQ ID NO: 10339. Partial BLAST results of SEQ ID NO: 10338 against GenBank are given below:

```
>gi|133593|sp|P18457|RRPB_CVPFS   RNA-DIRECTED RNA POLYMERASE (ORF1B)
 gi|93934|pir||A43489   RNA-directed RNA polymerase (EC 2.7.7.48) - porcine
transmissible gastroenteritis virus (fragment)
 gi|833161|emb|CAA37284.1|       polymerase  [Transmissible  gastroenteritis
virus]
          Length = 533

 Score =  131 bits (329), Expect = 3e-30
 Identities = 55/89 (61%), Positives = 69/89 (77%), Gaps = 1/89 (1%)

Query: 1   MLWCKDGHVETFYPKLQASQAWQPGVAMPNLYKMQRMLLEKCDLQNYGENAVIPKGIMMN 60
            MLWC++ H++TFYP+LQ+++ W PG +MP LYK+QRM LE+C+L NYG     +P GI  N
Sbjct: 217 MLWCENSHIKTFYPQLQSAE-WNPGYSMPTLYKIQRMCLERCNLYNYGAQVKLPDGITTN 275

Query: 61  VAKYTQLCQYLNTLTLAVPSNMRVIHFGA 89
            V KYTQLCQYLNT TL VP  MRV+H GA
Sbjct: 276 VVKYTQLCQYLNTTTLCVPHKMRVLHLGA 304
```

These results indicate that SEQ ID NO: 10338 has functional similarities to an RNA-directed RNA polymerase of porcine transmissible gastroenteritis virus.

Partial BLAST results of SEQ ID NO: 10339 against GenBank are given below:

```
>gb|AAL57305.1| replicase [bovine coronavirus]
          Length = 7094

 Score =  139 bits (351), Expect = 7e-33
 Identities = 64/108 (59%), Positives = 78/108 (72%)

Query: 1     MSVISKVVKVTIDYAEISFMLWCKDGHVETFYPKLQASQAWQPGVAMPNLYKMQRMLLEK 60
              M+ +SKVV V +D+ +  FMLWC D  V TFYP+LQA+  W+PG +MP LYK      +E+
Sbjct:   6760   LNCVSKVVNVNVDFKDFQFMLWCNDEKVMTFYPRLQAASDWKPGYSMPVLYKYLNSPMER
6819

Query: 61    CDLQNYGENAVIPKGIMMNVAKYTQLCQYLNTLTLAVPSNMRVIHFGA 108
                L NYG+    +P G MMNVAKYTQLCQYLNT TLAVP NMRV+H GA
Sbjct: 6820 VSLWNYGKPVTLPTGCMMNVAKYTQLCQYLNTTTLAVPVNMRVLHLGA 6867
```

These results indicate that SEQ ID NO: 10339 has functional similarities to a replicase of bovine coronavirus.

The SARS virus may contain polymorphism at the Glu-20 residue of SEQ ID NO: 10338. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 10338, wherein said polypeptide includes an amino acid sequence selected from the group consisting of ASQAW (SEQ ID NO: 10348) and ASRAW (SEQ ID NO: 10349). The invention includes a fragment of a polypeptide comprising SEQ ID NO: 10338, wherein said fragment includes an amino acid sequence selected from the group consisting of SEQ ID NO: 10348 and SEQ ID NO: 10349.

The SARS virus may contain polymorphism at the Ser-80 residue of SEQ ID NO: 10338. below. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 10338, wherein said polypeptide includes an amino acid sequence selected from the group consisting of VPSNM (SEQ ID NO: 10350) and VPTNM (SEQ ID NO: 10351). The invention includes a fragment of a polypeptide comprising SEQ ID NO: 10338, wherein said fragment includes an amino acid sequence selected from the group consisting of SEQ ID NO: 10350 and SEQ ID NO: 10351.

The invention also includes polynucleotide sequences which can be used as probes for diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention includes a polynucleotide sequence comprising one or more of the primer sequences identified in Table 32. The invention further includes polynucleotide sequence comprising the complement of one or more of the primer sequences identified in Table 32.

The invention provides a SARS polynucleotide sequence SEQ ID NO: 10505. All six reading frames of this sequence are shown in Figure 127 (see also Figure 132). The constituent amino acid sequences from Figure 127, having at least 4 amino acids, are listed as SEQ ID NOS: 10506 to 10570.

The invention includes a polynucleotide sequence comprising SEQ ID NO: 10505. The invention also provides polynucleotide sequences having sequence identity to SEQ ID NO: 10505. The invention also provides for polynucleotide sequences comprising fragments of SEQ ID NO: 10505. In one embodiment, the polynucleotide fragment does not consist entirely of a known SARS virus polynucleotide sequence or of a known coronavirus polynucleotide sequence.

The invention includes an amino acid sequence encoded by the polynucleotide sequence of SEQ ID NO: 10505, including the amino acid sequences shown in Figure 127, and particularly those selected from the group consisting of SEQ ID NO^S: 10506 to 10570. Preferably, the amino acid sequence comprises SEQ ID NO: 10532 and/or SEQ ID NO: 10533.

The invention also provides amino acid sequences having sequence identity to an amino acid sequence encoded by SEQ ID NO: 10505. The invention provides amino acid sequences

having sequence identity to an amino acid sequence selected from the group consisting of the sequences shown in Figure 127, and in particular SEQ ID NO^S: 10506 to 10570.

The invention also provides fragments of amino acid sequences encoded by SEQ ID NO: 10505. The invention also provides fragments of amino acid sequences selected from the group consisting of SEQ ID NO^S: 10506 to 10570. In one embodiment, the fragment does not consist entirely of a known amino acid sequence of a SARS virus or a known amino acid sequence of a coronavirus.

In one embodiment, the invention includes a polypeptide comprising an amino acid sequence from the 5'3' Frame 3 of Figure 127. Some encoded open reading frames within this translation are: SEQ ID NO: 10533; SEQ ID NO: 10571; SEQ ID NO: 10572; SEQ ID NO: 10573; SEQ ID NO: 10574.

The invention includes a polypeptide sequence comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 10533, SEQ ID NO: 10571, SEQ ID NO: 10572, SEQ ID NO: 10573 and SEQ ID NO: 10574. The invention includes a polypeptide having sequence identity to the amino acid sequence selected from the group consisting of SEQ ID NO: 10533, SEQ ID NO: 10571, SEQ ID NO: 10572, SEQ ID NO: 10573 and SEQ ID NO: 10574. The invention includes a fragment of a polypeptide sequence comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 10533, SEQ ID NO: 10571, SEQ ID NO: 10572, SEQ ID NO: 10573 and SEQ ID NO: 10574.

Partial BLAST results of SEQ ID NO: 10533 against GenBank are given below:

```
>gi|7739601|gb|AAF68926.1|AF207902_11        nucleocapsid   protein   [murine
hepatitis virus strain ML-11]
          Length = 451

 Score =  147 bits (370), Expect = 3e-34
 Identities = 102/252 (40%), Positives = 137/252 (54%), Gaps = 18/252 (7%)

Query: 49  SWFTALTQHGK-EELRFPRGQGVPINTNSGPDDQIGYYRRATRR-VRGGDGKMKELSPRW 106
           SWF+ +TQ  K +E +F +GQGVPI +      +Q GY+ R  RR  +  DG+ K+L PRW
Sbjct: 63  SWFSGITQFQKGKEFQFAQGQGVPIASGIPASEQKGYWYRHNRRSFKTPDGQHKQLLPRW 122

Query: 107 YFYYLGTGPEASLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLP 166
           YFYYLGTGP A   YG + EG+VWVA++ A       +  R+P+++ A   +   GT LP
Sbjct: 123 YFYYLGTGPHAGAEYGDDIEGVVWVASQQADTKTTADVVERDPSSHEAIPTRFAPGTVLP 182

Query: 167 KGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLN 226
           +GFY EGS   + AS   S        N    SS    PA          +A L+L +L
Sbjct: 183 QGFYVEGSGRSAPASRSGSRSQSRGPNNRARSSSNQRQPASAVKPDMAEEIAALVLAKLG 242

Query: 227 QLESKVSGKGQQQQGQTVTKKSAAEASK----KPRQKRTATKQYNVTQAFGRRGPEQTQG 282
           +         GQ +Q   VTK+SA E  +    KPRQKRT  KQ  V Q FG+RGP Q
Sbjct: 243 K------DAGQPKQ---VTKQSAKEVRQKILTKPRQKRTPNKQCPVQQCFGKRGPNQ--- 290

Query: 283 NFGDQDLIRQGT 294
           NFG  ++++ GT
Sbjct: 291 NFGGSEMLKLGT 302
```

```
>gi|3132999|gb|AAC16422.1|    nucleocapsid protein [murine hepatitis virus
strain 2]
          Length = 451

 Score =  147 bits (370), Expect = 3e-34
 Identities = 102/252 (40%), Positives = 137/252 (54%), Gaps = 18/252 (7%)

Query: 49  SWFTALTQHGK-EELRFPRGQGVPINTNSGPDDQIGYYRRATRR-VRGGDGKMKELSPRW 106
           SWF+ +TQ  K +E +F +GQGVPI +     +Q GY+ R  RR  +  DG+ K+L PRW
Sbjct: 63  SWFSGITQFQKGKEFQFAQGQGVPIASGIPASEQKGYWYRHNRRSFKTPDGQHKQLLPRW 122

Query: 107 YFYYLGTGPEASLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLP 166
           YFYYLGTGP A   YG + EG+VWVA++ A       +  R+P+++ A   +   GT LP
Sbjct: 123 YFYYLGTGPHAGAEYGDDIEGVVWVASQQADTKTTADVVERDPSSHEAIPTKFAPGTVLP 182

Query: 167 KGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLN 226
           +GFY EGS   + AS   S        N    SS    PA          +A L+L +L
Sbjct: 183 QGFYVEGSGKSAPASRSGSRSQSRGPNNRARSSSNQRQPASAVKPDMAEEIAALVLAKLG 242

Query: 227 QLESKVSGKGQQQQGQTVTKKSAAEASK----KPRQKRTATKQYNVTQAFGRRGPEQTQG 282
           +        GQ +Q   VTK+SA E  +    KPRQKRT  KQ  V Q FG+RGP Q
Sbjct: 243 K------DAGQPKQ---VTKQSAKEVRQKILTKPRQKRTPNKQCPVQQCFGKRGPNQ--- 290

Query: 283 NFGDQDLIRQGT 294
           NFG  ++++ GT
Sbjct: 291 NFGGSEMLKLGT 302


>gi|127877|sp|P03417|NCAP_CVMJH   Nucleocapsid protein
 gi|74859|pir||VHIHMJ     nucleocapsid protein - murine hepatitis virus
(strain JHM)
 gi|58973|emb|CAA25497.1|   nucleocapsid protein [Murine hepatitis virus]
          Length = 455

 Score =  146 bits (369), Expect = 4e-34
 Identities = 110/254 (43%), Positives = 142/254 (55%), Gaps = 22/254 (8%)

Query: 49  SWFTALTQHGK-EELRFPRGQGVPINTNSGPDDQIGYYRRATRR-VRGGDGKMKELSPRW 106
           SWF+ +TQ  K +E +F +GQGVPI        Q GY+ R  RR  +  DG+ K+L PRW
Sbjct: 67  SWFSGITQFQKGKEFQFAQGQGVPIANGIPASQQKGYWYRHNRRSFKTPDGQQKQLLPRW 126

Query: 107 YFYYLGTGPEASLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLP 166
           YFYYLGTGP A   YG + EG+VWVA++ A       I  R+P+++ A   +   GT LP
Sbjct: 127 YFYYLGTGPYAGAEYGDDIEGVVWVASQQAETRTSADIVERDPSSHEAIPTRFAPGTVLP 186

Query: 167 KGFYAEGSRGGSQASSRSSSR--SRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDR 224
           +GFY EGS G S  +SRS SR  SRG   N    SS    PA          +A L+L +
Sbjct: 187 QGFYVEGS-GRSAPASRSGSRPQSRG-PNNRARSSSNQRQPASTVKPDMAEEIAALVLAK 244

Query: 225 LNQLESKVSGKGQQQQGQTVTKKSAAEASK----KPRQKRTATKQYNVTQAFGRRGPEQT 280
           L +        GQ +Q   VTK+SA E  +    KPRQKRT  KQ  V Q FG+RGP Q
Sbjct: 245 LGK------DAGQPKQ---VTKQSAKEVRQKILNKPRQKRTPNKQCPVQQCFGKRGPNQ- 294

Query: 281 QGNFGDQDLIRQGT 294
             NFG  ++++ GT
Sbjct: 295 --NFGGPEMLKLGT 306
```

```
>gi|6625766|gb|AAF19389.1|AF201929_7       nucleocapsid   protein   [murine
hepatitis virus strain 2]
 gi|7769348|gb|AAF69338.1|AF208066_11       nucleocapsid   protein   [murine
hepatitis virus]
          Length = 451

 Score =  146 bits (368), Expect = 5e-34
 Identities = 102/252 (40%), Positives = 137/252 (54%), Gaps = 18/252 (7%)

Query: 49  SWFTALTQHGK-EELRFPRGQGVPINTNSGPDDQIGYYRRATRR-VRGGDGKMKELSPRW 106
           SWF+ +TQ  K +E +F +GQGVPI +     +Q GY+ R  RR  +  DG+ K+L PRW
Sbjct: 63  SWFSGITQFQKGKEFQFAQGQGVPIASGIPASEQKGYWYRHNRRSFKTPDGQHKQLLPRW 122

Query: 107 YFYYLGTGPEASLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLP 166
           YFYYLGTGP A   YG + EG+VWVA++ A       +  R+P+++ A   +   GT LP
Sbjct: 123 YFYYLGTGPHAGAEYGDDIEGVVWVASQQADTKTTADVVERDPSSHEAIPTRFAPGTVLP 182

Query: 167 KGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLN 226
           +GFY EGS   + AS   S        N    SS    PA          +A L+L +L
Sbjct: 183 QGFYVEGSGRSAPASRSGSRSQSRGPNNRARSSSNQRQPASAVKPDMAEEIAALVLAKLG 242

Query: 227 QLESKVSGKGQQQQGQTVTKKSAAEASK----KPRQKRTATKQYNVTQAFGRRGPEQTQG 282
           +      GQ +Q   VTK+SA E  +    KPRQKRT  KQ  V Q FG+RGP Q
Sbjct: 243 K------DAGQPKQ---VTKQSAKEVRQKILTKPRQKRTPNKQCPVQQCFGKRGPNQ--- 290

Query: 283 NFGDQDLIRQGT 294
           NFG  ++++ GT
Sbjct: 291 NFGGSEMLKLGT 302


>gi|21734854|gb|AAM77005.1|AF481863_7   phosphorylated nucleocapsid protein
N [porcine hemagglutinating encephalomyelitis virus]
          Length = 449

 Score =  145 bits (366), Expect = 8e-34
 Identities = 107/253 (42%), Positives = 145/253 (57%), Gaps = 18/253 (7%)

Query: 49  SWFTALTQHGK-EELRFPRGQGVPINTNSGPDDQIGYYRRATRR-VRGGDGKMKELSPRW 106
           SWF+ +TQ  K +E  F  GQGVPI       +  GY+ R  RR  +  DG  ++L PRW
Sbjct: 64  SWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQLLPRW 123

Query: 107 YFYYLGTGPEASLPYGANKEGIVWVATEGA-LNTPKDHIGTRNPNNNAATVLQLPQGTTL 165
           YFYYLGTGP A   YG + +G+ WVA+  A +NTP D I  R+P+++ A   + P GT L
Sbjct: 124 YFYYLGTGPHAKHQYGTDIDGVFWVASNQADINTPAD-IVDRDPSSDEAIPTRFPPGTVL 182

Query: 166 PKGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRL 225
           P+G+Y EGS G S  +SRS+SR+  N   S    SR NS  R ++ G    +A    D++
Sbjct: 183 PQGYYIEGS-GRSAPNSRSTSRA-PNRAPSAGSRSRANSGNRTSTPGVTPDMA----DQI 236

Query: 226 NQLESKVSGKGQQQQGQTVTKKSAAEASK----KPRQKRTATKQYNVTQAFGRRGPEQTQ 281
             L     GK    + Q VTK++A E  +    KPRQKR+  KQ  V Q FG+RGP Q
Sbjct: 237 ASLVLAKLGK-DATKPQQVTKQTAKEVRQKILNKPRQKRSPNKQCTVQQCFGKRGPNQ-- 293

Query: 282 GNFGDQDLIRQGT 294
            NFG  ++++ GT
Sbjct: 294 -NFGGGEMLKLGT 305
```

```
>gi|23295765|gb|AAL80036.1|              nucleocapsid    protein    [porcine
hemagglutinating encephalomyelitis virus]
          Length = 449

 Score =  145 bits (365), Expect = 1e-33
 Identities = 107/253 (42%), Positives = 145/253 (57%), Gaps = 18/253 (7%)

Query: 49  SWFTALTQHGK-EELRFPRGQGVPINTNSGPDDQIGYYRRATRR-VRGGDGKMKELSPRW 106
           SWF+ +TQ  K +E  F  GQGVPI       +  GY+ R  RR  +  DG  ++L PRW
Sbjct: 64  SWFSGITQFQKGKEFEFAEGQGVPIAPGVPSTEAKGYWYRHNRRSFKTADGNQRQLLPRW 123

Query: 107 YFYYLGTGPEASLPYGANKEGIVWVATEGA-LNTPKDHIGTRNPNNNAATVLQLPQGTTL 165
           YFYYLGTGP A   YG + +G+ WVA+  A +NTP D I  R+P+++ A   + P GT L
Sbjct: 124 YFYYLGTGPHAKDQYGTDIDGVFWVASNQADINTPAD-IVDRDPSSDEAIPTRFPPGTVL 182

Query: 166 PKGFYAEGSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRL 225
           P+G+Y EGS G S  +SRS+SR+  N   S   SR NS R ++ G    +A    D++
Sbjct: 183 PQGYYIEGS-GRSAPNSRSTSRA-PNRAPSAGSRSRANSGNRTSTPGVTPDMA----DQI 236

Query: 226 NQLESKVSGKGQQQQGQTVTKKSAAEASK----KPRQKRTATKQYNVTQAFGRRGPEQTQ 281
             L    GK   + Q VTK++A E  +    KPRQKR+  KQ  V Q FG+RGP Q
Sbjct: 237 ASLVLAKLGK-DATKPQQVTKQTAKEVRQKILNKPRQKRSPNKQCTVQQCFGKRGPNQ-- 293

Query: 282 GNFGDQDLIRQGT 294
           NFG  ++++ GT
Sbjct: 294 -NFGGGEMLKLGT 305
```

These results indicate that SEQ ID NO: 10533 has functional similarities to a coronavirus nucleocapsid protein.

In one embodiment, the invention comprises an amino acid sequence from the 5'3' Frame 1 of Figure 127 *e.g.* SEQ ID NO^S: 10506-10514. Some encoded open reading frames within this region are SEQ ID NO^S: 10575 to 10578.

Accordingly, the invention includes a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 10575, SEQ ID NO: 10576, SEQ ID NO: 10577 and SEQ ID NO: 10578. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to a sequence selected from the group consisting of SEQ ID NO: 10097, SEQ ID NO: 10576, SEQ ID NO: 10577 and SEQ ID NO: 10578. The invention includes a fragment of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 10097, SEQ ID NO: 10576, SEQ ID NO: 10577 and SEQ ID NO: 10578.

In one embodiment, the invention includes a polypeptide comprising an amino acid sequence from the 3'5' Frame 2 of Figure 127 *e.g.* SEQ ID NO^S: 10547-10559. An open reading frame within this region is SEQ ID NO: 10579.

The invention includes a polypeptide comprising an amino acid sequence of SEQ ID NO: 10579. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 10579. The invention includes a fragment of a polypeptide comprising an amino acid sequence of SEQ ID NO: 10579.

The invention also includes polynucleotide sequences which can be used as probes for diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention includes a polynucleotide sequence comprising one or more of the primer sequences identified in Table 33. The invention further includes polynucleotide sequence comprising the complement of one or more of the primer sequences identified in Table 33.

The invention includes a polynucleotide sequence comprising SEQ ID NO: 11323. A polypeptide encoded by SEQ ID NO: 11323 is SEQ ID NO: 11324.

The invention includes a polypeptide comprising SEQ ID NO: 11324, sequence having sequence identity to SEQ ID NO: 11324 and fragments of SEQ ID NO: 11324. The invention includes a fragment of SEQ ID NO: 11324, wherein said polypeptide fragment begins with a Methionine.

Accordingly, the invention includes a polynucleotide sequence comprising SEQ ID NO: 11323. It also provides polynucleotide sequences having sequence identity to SEQ ID NO: 11323. The invention also provides for polynucleotide sequences comprising fragments of SEQ ID NO: 11323. In one embodiment, the polynucleotide fragment does not consist entirely of a known SARS polynucleotide sequence or a known coronavirus polynucleotide sequence.

The invention includes an amino acid sequence encoded by the polynucleotide sequence SEQ ID NO: 11323, including the amino acid sequence of SEQ ID NO: 11324.

The invention also provides amino acid sequences having sequence identity to an amino acid sequence encoded by SEQ ID NO: 11323. The invention provides amino acid sequences having sequence identity to SEQ ID NO: 11324.

The invention provides fragments of amino acid sequences encoded by SEQ ID NO: 11323. The invention also provides fragments of amino acid sequences of SEQ ID NO: 11324. In one embodiment, the fragment does not consist entirely of a known SARS amino acid sequence or a known coronavirus amino acid sequence.

The invention also includes polynucleotide sequences which can be used as probes for diagnostic reagents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample. The invention includes a polynucleotide sequence comprising one or more of the primer sequences identified as SEQ ID NO^S: 11325-11440 (left part) and SEQ ID NO^S: 11441-11551 (right part). The invention further includes polynucleotide sequence comprising the complement of one or more of the primer sequences identified as SEQ ID NO^S: 11325-11551.

The invention includes a polypeptide comprising SEQ ID NO: 11552. The SARS virus contains polymorphism at the Isoleucine residue Ile-324. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 11552, wherein

said polypeptide includes an amino acid sequence selected from the group consisting of YSYAI (SEQ ID NO: 11553), SYAIH (SEQ ID NO: 11554), YAIHH (SEQ ID NO: 11555), IHHDK (SEQ ID NO: 11556), SYAI (SEQ ID NO: 11557), YAIH (SEQ ID NO: 11558), AIHH (SEQ ID NO: 11559), IHHD (SEQ ID NO: 11560), YAI, AIH, and IHH. The invention includes a fragment of a polypeptide comprising SEQ ID NO: 11552, wherein said fragment includes an amino acid sequence selected from the group consisting of YSYAI (SEQ ID NO: 11553), SYAIH (SEQ ID NO: 11554), YAIHH (SEQ ID NO: 11555), IHHDK (SEQ ID NO: 11556), SYAI (SEQ ID NO: 11557), YAIH (SEQ ID NO: 11558), AIHH (SEQ ID NO: 11559), IHHD (SEQ ID NO: 11560), YAI, AIH, and IHH.

The invention includes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 11561 and SEQ ID NO: 11562. The invention includes a fragment of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 11561 and SEQ ID NO: 11562.

The invention includes a diagnostic kit comprising a polypeptide comprising at least one of the amino acid sequences selected from the group consisting of SEQ ID NO^S: 11561 and 11562. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding a polypeptide comprising at least one of the amino acid sequences selected from the group consisting of SEQ ID NO^S: 11561 and 11562. The invention includes an immunogenic composition comprising a polypeptide comprising at least one of the amino acid sequences selected from the group consisting of SEQ ID NO^S: 11561 and 11562. The invention includes an antibody which recognizes a polypeptide comprising at least one of the amino acid sequences selected from the group consisting of SEQ ID NO^S: 11561 and 11562.

The invention includes a polynucleotide sequence SEQ ID NO: 11563 or a fragment thereof or a sequence having sequence identity thereto. Polypeptide sequences which can be translated from SEQ ID NO: 11563 are shown in Figure 128. The constituent amino acid sequences from Figure 128, having at least 4 amino acids, are listed as SEQ ID NO^S: 11564 to 11617.

The invention includes a polypeptide sequence selected from the group consisting of the sequences of Figure 128, or a fragment thereof or a sequence having sequence identity thereto *e.g.* SEQ ID NO^S: 11563 to 11617.

A polypeptide sequence within SEQ ID NO: 11600 is SEQ ID NO: 11618. The invention includes a polypeptide comprising SEQ ID NO: 11618, or a fragment thereof or a sequence having sequence identity thereto.

A polypeptide sequence within SEQ ID NO: 11602 is SEQ ID NO: 11641. The invention includes a polypeptide comprising SEQ ID NO: 11641, or a fragment thereof or a sequence having sequence identity thereto.

A polypeptide sequence within SEQ ID NO: 11609 is SEQ ID NO: 11619.

The invention includes a polynucleotide encoding (i) an amino acid sequence selected from the group consisting of: (1) the amino acid sequences of Figure 128, and in particular SEQ ID NO^S: 11564-11617; (2) SEQ ID NO: 11618; and (3) SEQ ID NO: 11619, or (ii) a fragment thereof. The invention includes a diagnostic kit comprising a one or more of these proteins. The invention includes a diagnostic kit comprising a polynucleotide sequence encoding one or more of these polypeptide sequences. The invention includes an antibody which recognizes one or more of the polypeptide sequences.

The SARS virus may contain polymorphism at isoleucine residue Ile-326 in SEQ ID NO: 11620 (Chi-PEP3). The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 11620, wherein said polypeptide includes an amino acid sequence selected from the group consisting of YA**I**HH (SEQ ID NO: 11621) and YA**T**HH (SEQ ID NO: 11622). The invention includes a fragment of a polypeptide comprising SEQ ID NO: 11620, wherein said fragment includes an amino acid sequence selected from the group consisting of YA**I**HH (SEQ ID NO: 11621) and YA**T**HH (SEQ ID NO: 11622).

The SARS virus may contain polymorphism at glutamine residue Gln-830 in SEQ ID NO: 11620. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 11620, wherein said polypeptide includes an amino acid sequence selected from the group consisting of AS**Q**AW (SEQ ID NO: 11623) and AS**R**AW (SEQ ID NO: 11624). The invention includes a fragment of a polypeptide comprising SEQ ID NO: 11620, wherein said fragment includes an amino acid sequence selected from the group consisting of AS**Q**AW (SEQ ID NO: 11623) and AS**R**AW (SEQ ID NO: 11624).

The SARS virus may contain polymorphism at aspartic acid residue Asp-935 in SEQ ID NO: 11620. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 11620, wherein said polypeptide includes an amino acid sequence selected from the group consisting of DA**D**ST (SEQ ID NO: 11625) and DA**Y**ST (SEQ ID NO: 11626). The invention includes a fragment of a polypeptide comprising SEQ ID NO: 11620, wherein said fragment includes an amino acid sequence selected from the group consisting of DA**D**ST (SEQ ID NO: 11625) and DA**Y**ST (SEQ ID NO: 11626).

The SARS virus may contain polymorphism at serine residue Ser-577 in SEQ ID NO: 11627 (Chi-PEP4). The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 11627, wherein said polypeptide includes an amino

acid sequence selected from the group consisting of PCSFG (SEQ ID NO: 11628) and PCAFG (SEQ ID NO: 11629). The invention includes a fragment of a polypeptide comprising SEQ ID NO: 11627, wherein said fragment includes an amino acid sequence selected from the group consisting of PCSFG (SEQ ID NO: 11628) and PCAFG (SEQ ID NO: 11629).

The SARS virus may contain polymorphism at valine residue Val-68 in SEQ ID NO: 11630 (Chi-PEP8). The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 11630, wherein said polypeptide includes an amino acid sequence selected from the group consisting of LAVVY (SEQ ID NO: 11631) and LAAVY (SEQ ID NO: 11632). The invention includes a fragment of a polypeptide comprising SEQ ID NO: 11630, wherein said fragment includes an amino acid sequence selected from the group consisting of LAVVY (SEQ ID NO: 11631) and LAAVY (SEQ ID NO: 11632).

The SARS virus may contain polymorphism at isoleucine residue Ile-50 in SEQ ID NO: 11633 (Chi-PEP13). The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 11633, wherein said polypeptide includes an amino acid sequence selected from the group consisting of NNIAS (SEQ ID NO: 11634) and NNTAS (SEQ ID NO: 11635). The invention includes a fragment of a polypeptide comprising SEQ ID NO: 11633, wherein said fragment includes an amino acid sequence selected from the group consisting of NNIAS (SEQ ID NO: 11634) and NNTAS (SEQ ID NO: 11635).

The SARS virus may contain a polymorphism at Serine residue Ser-943 in SEQ ID NO: 11636. The invention includes a polypeptide comprising an amino acid sequence having sequence identity to SEQ ID NO: 11636, wherein said polypeptide includes an amino acid sequence selected from the group consisting of AVSAC (SEQ ID NO: 11637) and AVGAC (SEQ ID NO: 11638). The invention includes a fragment of a polypeptide comprising SEQ ID NO: 11636, wherein said fragment includes an amino acid seuence selected from the group consisting of AVSAC (SEQ ID NO: 11637) and AVGAC (SEQ ID NO: 11638).

The invention includes a polynucleotide SEQ ID NO: 11639, or a fragment thereof or a sequence having sequence identity thereto. The invention includes a polypeptide encoded by the polynucleotide sequence set forth in SEQ ID NO: 11639, or a fragment thereof or a polypeptide sequence having sequence identity thereto.

The invention includes a polynucleotide set forth in SEQ ID NO: 11640, or a fragment thereof or a sequence having sequence identity thereto. The invention includes a polypeptide encoded by the polynucleotide sequence set forth in SEQ ID NO: 11640, or a fragment thereof or a polypeptide sequence having sequence identity thereto.

The invention includes each of the polynucleotides identified above. The invention includes each of the polynucleotides set forth in the sequence listing. The invention further

includes polynucleotides having sequence identity to each of the polynucleotides identified above. The degree of sequence identity is preferably greater than 50% (*e.g.*, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more).

The invention includes polynucleotide sequences comprising fragments of each of the polynucleotide sequences identified above. The fragments should comprise at least *n* consecutive polynucleotides from a particular SEQ ID NO:, and, depending on the sequence, *n* is 7 or more (*e.g.*, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 80, 90, 100, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300 or more).

The invention includes each of the amino acid sequences encoded by each of the polynucleotide sequences identified above. The invention includes each of the amino acid sequences encoded by each of the polynucleotide sequences set forth in the sequence listing. The invention further includes amino acid sequences having sequence identity to the amino acid sequences encoded by each of the polynucleotide sequences identified above. The degree of sequence identity is preferably greater than 50% (*e.g.*, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more). The invention further includes fragments of amino acid sequences encoded by each of the polynucleotide sequences identified above. The fragments should comprise at least *n* consecutive amino acids from a particular SEQ ID NO:, and, depending on the sequence, *n* is 7 or more (*e.g.*, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 80, 90, 100, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300 or more).

The invention includes each of the amino acid sequences identified above. The invention includes each of the amino acid sequence set forth in the sequence listing. The invention further includes amino acid sequences having sequence identity to each of the amino acid sequences identified above. The degree of sequence identity is preferably greater than 50% (*e.g.*, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more).

The invention further includes fragments of the amino acid sequences identified above. The fragments should comprise at least *n* consecutive amino acids from a particular SEQ ID NO:, and, depending on the sequence, *n* is 7 or more (*e.g.*, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55,

60, 65, 70, 80, 90, 100, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300 or more).

The invention includes polynucleotides encoding each of the amino acid sequences identified above. The invention includes polynucleotides encoding each of the amino acid sequences set forth in the sequence listing. The invention further includes polynucleotides having sequence identity with each of the polynucleotides encoding each of the amino acid sequences identified above. The degree of sequence identity is preferably greater than 50% (*e.g.*, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more).

The invention further includes fragments of polynucleotides encoding each of the amino acid sequences identified above. The fragments should comprise at least *n* consecutive polynucleotides from a particular SEQ ID NO:, and, depending on the sequence, *n* is 7 or more (*e.g.*, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 80, 90, 100, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300 or more).

As described in more detail below, polynucleotides for use as primers and/or as probes may contain at least 4 or 8 contiguous nucleotides from a polynucleotide sequence of the invention *e.g.* at least 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 contiguous nucleotides and up to about 50, 75, 100, 200 contiguous nucleotides or more. While 6-8 nucleotides may be a workable length, sequences of 10-12 nucleotides are preferred, and about 13, 14, 15, 16, 17, 18, 19, 20, or 21 or more nucleotides or more appears optimal for hybridisation.

In one embodiment, the invention is directed to polynucleotides and amino acid sequences that do not consist entirely of a known SARS virus polynucleotide or amino acid sequence or of a known coronavirus polynucleotide or amino acid sequence. In one embodiment, the polynucleotides and amino acid sequences of the invention do not consist entirely of the sequence SEQ ID NO: 1. In another embodiment, the polynucleotides and amino acid sequences of the invention do not consist entirely of the sequence SEQ ID NO: 2. SEQ ID NO: 9967 is a SARS genome sequence of the Frankfurt (FRA) isolate (GenBank: AY310120). Compared to SEQ ID NO: 1, it differs at nucleotides 2546, 2590, 11437, 18954, 19073, 20585, 20899, 23209, 24922, 26589 & 28257; compared to SEQ ID NO:2, it differs at nucleotides 2560, 7922, 11451, 16625, 18968 & 19067. Further genome sequences have become available from GenBank, since this application was originally filed, under accession numbers including AY559097, AY559096, AY559095, AY559094, AY559093, AY559092, AY559091, AY559090, AY559089, AY559088, AY559087, AY559086, AY559085, AY559084, AY559083, AY559082, AY559081, AY274119,

AY323977, AY291315, AY502932, AY502931, AY502930, AY502929, AY502928, AY502927, AY502926, AY502925, AY502924, AY502923, AY291451, AY390556, AY395003, AY395002, AY395001, AY395000, AY394999, AY394998, AY394997, AY394996, AY394995, AY394994, AY394993, AY394992, AY394991, AY394990, AY394989, AY394987, AY394986, AY394985, AY394983, AY394979, AY394978, AY508724, AY394850, AY463059, AY463060, AY313906, AY310120, AY461660, AY485278, AY485277, AY345988, AY345987, AY345986, AY282752, AY357076, AY357075, AY350750, AY304495, AY304488, AY304486, AY427439, AY283798, AY278491, AY278489, AY362699, AY362698, AY283797, AY283796, AY283795, AY283794, AY278741, AY351680, AP006561, AP006560, AP006559, AP006558, AP006557, AY278554, AY348314, AY338175, AY338174, AY321118, AY279354, AY278490, AY278487, AY297028, AY278488, and NC_004718.

In another embodiment, the invention is directed to polynucleotides that encode proteins which are not immunologically cross reactive with a protein of a mouse hepatitis virus, a bovine coronavirus or an avian infectious bronchitis virus. In another embodiment, the invention is directed to proteins which are not immunologically cross reactive with a protein of a mouse hepatitis virus, a bovine coronavirus or an avian infectious bronchitis virus.

Each of the polynucleotides identified above may be used to encode a portion of a fusion protein. Accordingly, the invention compries one or more of the polynucleotides identified above wherein the polynucleotides encoding for the start codon are removed. The invention further comprises one or more of the amino acids identified above wherein the starting methionine is removed.

Any of the polynucleotide or amino acid sequences discussed above may be used in vaccines for the treatment or prevention of SARS virus infection, including as a SARS viral antigen. Additionally, any of the polynucleotides or amino acid sequences discussed above may be used as diagnostic reagents, or in kits (comprising such reagents) or in methods used to diagnose or identify the presence or absence of a SARS virus in a biological sample.

SARS viral antigens of the invention may include a polypeptide with 99%, 95%, 90%, 85%, or 80% homology to one or more of the group consisting of the following proteins: nonstructural protein 2 (NS2); hemagglutinin-esterase glycoprotein (HE) (also referred to as E3), spike glycoprotein (S) (also referred to as E2), nonstructural region 4 (NS4), envelope (small membrane) protein (E) (also referred to as sM), membrane glycoprotein (M) (also referred to as E1), nucleocapsid phosphoprotein (N) or RNA dependent RNA polymerase (pol).

A detailed discussion of Coroavirus biology can be found in *Fields Virology* (2nd ed), Fields *et al.* (eds.), B.N. Raven Press, New York, NY., Chapter 35.

Another example of a SARS virus isolate is set forth in Example 1 below. The invention includes each of the polypeptide and polynucleotide sequences identified in Example 1. In

addition, the invention includes vaccine formulations comprising one or more of the polypeptide or polynucleotide sequences identified in Example 1. The invention includes diagnostic regaents, kits (comprising such reagents) and methods which can be used to diagnose or identify the presence or absence of a SARS virus in a biological sample using one or more of the polypeptide or polynucleotide sequences identified in Example 1. The invention includes methods for the treatment or prevention of SARS virus infection utilizing small molecule viral inhibitors and combinations of small molecule viral inhibitors and kits for the treatment of SARS. The small molecule inhibitors may specifically target one or more of the polypeptides or polynucleotides identified in Example 1.

Further discussion of terms used in the application follows below.

"Respiratory Virus" as used herein refers to a virus capable of infecting the human respiratory tract. Respiratory Viral Antigens suitable for use in the invention include Severe Acute Respiratory Syndrome virus, coronavirus, influenza virus, human rhinovirus (HRV), parainfluenza virus (PIV), respiratory syncytial virus (RSV), adenovirus, metapneumovirus, and rhinovirus.

The terms "polypeptide", "protein" and "amino acid sequence" as used herein generally refer to a polymer of amino acid residues and are not limited to a minimum length of the product. Thus, peptides, oligopeptides, dimers, mulimers, and the like, are included within the definition. Both full-length proteins and fragments thereof are encompassed by the definition. Minimum fragments of polypeptides useful in the invention can be at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or even 15 amino acids. Typically, polypeptides useful in this invention can have a maximum length suitable for the intended application. Generally, the maximum length is not critical and can easily be selected by one skilled in the art.

Polypeptides of the invention can be prepared in many ways *e.g.* by chemical synthesis (at least in part), by digesting longer polypeptides using proteases, by translation from RNA, by purification from cell culture (*e.g.* from recombinant expression), from the organism itself (*e.g.* after viral culture, or direct from patients), from a cell line source *etc.* A preferred method for production of peptides <40 amino acids long involves *in vitro* chemical synthesis (Bodanszky (1993) *Principles of Peptide Synthesis* (ISBN: 0387564314); Fields *et al.* (1997) *Methods in Enzymology 289: Solid-Phase Peptide Synthesis.* ISBN: 0121821900). Solid-phase peptide synthesis is particularly preferred, such as methods based on t-Boc or Fmoc (Chan & White (2000) *Fmoc Solid Phase Peptide Synthesis* ISBN: 0199637245) chemistry. Enzymatic synthesis (Kullmann (1987) *Enzymatic Peptide Synthesis.* ISBN: 0849368413) may also be used in part or in full. As an alternative to chemical synthesis, biological synthesis may be used *e.g.* the polypeptides may be produced by translation. This may be carried out in vitro or in vivo. Biological methods are in general restricted to the production of polypeptides based on L-amino

acids, but manipulation of translation machinery (*e.g.* of aminoacyl tRNA molecules) can be used to allow the introduction of D-amino acids (or of other non natural amino acids, such as iodotyrosine or methylphenylalanine, azidohomoalanine, *etc.*) (Ibba (1996) *Biotechnol Genet Eng Rev* 13:197-216.). Where D-amino acids are included, however, it is preferred to use chemical synthesis. Polypeptides of the invention may have covalent modifications at the C-terminus and/or N-terminus, particularly where they are for *in vivo* administration *e.g* by attachment of acetyl or carboxamide, as in the Fuzeon™ product.

Reference to polypeptides and the like also includes derivatives of the amino acid sequences of the invention. Such derivatives can include postexpression modifications of the polypeptide, for example, glycosylation, acetylation, phosphorylation, and the like. Amino acid derivatives can also include modifications to the native sequence, such as deletions, additions and substitutions (generally conservative in nature), so long as the protein maintains the desired activity. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the proteins or errors due to PCR amplification. Furthermore, modifications may be made that have one or more of the following effects: reducing toxicity; facilitating cell processing (*e.g.*, secretion, antigen presentation, *etc.*); and facilitating presentation to B-cells and/or T-cells.

"Fragment" or "Portion" as used herein refers to a polypeptide consisting of only a part of the intact full-length polypeptide sequence and structure as found in nature. For instance, a fragment can include a C-terminal deletion and/or an N-terminal deletion of a protein.

A "recombinant" protein is a protein which has been prepared by recombinant DNA techniques as described herein. In general, the gene of interest is cloned and then expressed in transformed organisms, as described further below. The host organism expressed the foreign gene to produce the protein under expression conditions.

The term "polynucleotide", as known in the art, generally refers to a nucleic acid molecule. A "polynucleotide" can include both double- and single-stranded sequences and refers to, but is not limited to, cDNA from viral, prokaryotic or eukaryotic mRNA, genomic RNA and DNA sequences from viral (*e.g.* RNA and DNA viruses and retroviruses) or prokaryotic DNA, and especially synthetic DNA sequences. The term also captures sequences that include any of the known base analogs of DNA and RNA, and includes modifications such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the nucleic acid molecule encodes a therapeutic or antigenic protein. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts that produce the antigens. Modifications of polynucleotides may have any number of effects including, for example, facilitating expression of the polypeptide product in a host cell.

Polynucleotides of the invention may be prepared in many ways *e.g.* by chemical synthesis (*e.g.* phosphoramidite synthesis of DNA) in whole or in part, by digesting longer nucleic acids using nucleases (*e.g.* restriction enzymes), by joining shorter nucleic acids or nucleotides (*e.g.* using ligases or polymerases), from genomic or cDNA libraries, *etc.*

A polynucleotide can encode a biologically active (*e.g.*, immunogenic or therapeutic) protein or polypeptide. Depending on the nature of the polypeptide encoded by the polynucleotide, a polynucleotide can include as little as 10 nucleotides, *e.g.*, where the polynucleotide encodes an antigen.

By "isolated" is meant, when referring to a polynucleotide or a polypeptide, that the indicated molecule is separate and discrete from the whole organism with which the molecule is found in nature or, when the polynucleotide or polypeptide is not found in nature, is sufficiently free of other biological macromolecules so that the polynucleotide or polypeptide can be used for its intended purpose. The polynucleotides and polypeptides of the invention are preferably isolated polynucleotides and isolated polypeptides.

"Antibody" as known in the art includes one or more biological moieties that, through chemical or physical means, can bind to or associate with an epitope of a polypeptide of interest. The antibodies of the invention include antibodies which specifically bind to a SARS viral antigen. The term "antibody" includes antibodies obtained from both polyclonal and monoclonal preparations, as well as the following: hybrid (chimeric) antibody molecules (see, for example, Winter *et al.* (1991) *Nature* 349: 293-299; and US Patent No. 4,816,567; $F(ab')_2$ and F(ab) fragments; F_v molecules (non-covalent heterodimers, see, for example, Inbar *et al.* (1972) *Proc Natl Acad Sci USA* 69:2659-2662; and Ehrlich *et al.* (1980) *Biochem* 19:4091-4096); single-chain Fv molecules (sFv) (see, for example, Huston *et al.* (1988) *Proc Natl Acad Sci USA* 85:5897-5883); dimeric and trimeric antibody fragment constructs; minibodies (see, *e.g.*, Pack *et al.* (1992) *Biochem* 31:1579-1584; Cumber *et al.* (1992) *J Immunology* 149B: 120-126); humanized antibody molecules (see, for example, Riechmann *et al.* (1988) *Nature* 332:323-327; Verhoeyan *et al.* (1988) *Science* 239:1534-1536; and U.K. Patent Publication No. GB 2,276,169, published 21 September 1994); and, any functional fragments obtained from such molecules, wherein such fragments retain immunological binding properties of the parent antibody molecule. The term "antibody" further includes antibodies obtained through non-conventional processes, such as phage display.

As used herein, the term "monoclonal antibody" refers to an antibody composition having a homogeneous antibody population. The term is not limited regarding the species or source of the antibody, nor is it intended to be limited by the manner in which it is made. Thus, the term encompasses antibodies obtained from murine hybridomas, as well as human monoclonal

antibodies obtained using human rather than murine hybridomas. See, *e.g.*, Cote, *et al.* *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, p 77.

An "immunogenic composition" as used herein refers to a composition that comprises an antigenic molecule where administration of the composition to a subject results in the development in the subject of a humoral and/or a cellular immune response to the antigenic molecule of interest. The immunogenic composition can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal or any other parenteral, mucosal or transdermal (*e.g.*, intra-rectally or intra-vaginally) route of administration.

The term "derived from" is used to identify the source of molecule (*e.g.*, a molecule can be derived from a polynucleotide, polypeptide, an immortalized cell line can be derived from any tissue, *etc.*). A first polynucleotide is "derived from" a second polynucleotide if it has the same or substantially the same basepair sequence as a region of the second polynucleotide, its cDNA, complements thereof, or if it displays sequence identity as described above. Thus, a first polynucleotide sequence is "derived from" a second sequence if it has (i) the same or substantially the same sequence as the second sequence or (ii) displays sequence identity to polypeptides of that sequence.

A first polypeptide is "derived from" a second polypeptide if it is (i) encoded by a first polynucleotide derived from a second polynucleotide, or (ii) displays sequence identity to the second polypeptides as described above. Thus, a polypeptide (protein) is "derived from" a particular SARS virus if it is (i) encoded by an open reading frame of a polynucleotide of that SARS virus, or (ii) displays sequence identity, as described above, to polypeptides of that SARS virus.

Both polynucleotide and polypeptide molecules can be physically derived from a SARS virus or produced recombinantly or synthetically, for example, based on known sequences.

A cultured cell or cell line is "derived from" another cell, cells or tissue if it is originally obtained from existing cells or tissue. Non-limiting examples of tissue that cells may be derived from include skin, retina, liver, kidney, heart, brain, muscle, intestinal, ovary, breast, prostate, cancerous tissue, tissue infected with one or more pathogens (*e.g.*, viruses, bacteria *etc.*) and the like. The cells described herein may also be derived from other cells including, but not limited to, primary cultures, existing immortalized cells line and/or other isolated cells.

An "antigen" refers to a molecule containing one or more epitopes (either linear, conformational or both) that will stimulate a host's immune system to make a humoral and/or cellular antigen-specific response. The term is used interchangeably with the term "immunogen." Normally, an epitope will include between about 3-15, generally about 5-15 amino acids. A B-cell epitope is normally about 5 amino acids but can be as small as 3-4 amino acids. A T-cell epitope, such as a CTL epitope, will include at least about 7-9 amino acids, and a

helper T-cell epitope at least about 12-20 amino acids. Normally, an epitope will include between about 7 and 15 amino acids, such as, 9, 10, 12 or 15 amino acids. The term "antigen" denotes both subunit antigens, (*i.e.*, antigens which are separate and discrete from a whole organism with which the antigen is associated in nature), as well as, killed, attenuated or inactivated bacteria, viruses, fungi, parasites or other microbes as well as tumor antigens, including extracellular domains of cell surface receptors and intracellular portions that may contain T-cell epitopes. Antibodies such as anti-idiotype antibodies, or fragments thereof, and synthetic peptide mimotopes, which can mimic an antigen or antigenic determinant, are also captured under the definition of antigen as used herein. Similarly, an oligonucleotide or polynucleotide that expresses an antigen or antigenic determinant *in vivo*, such as in gene therapy and DNA immunization applications, is also included in the definition of antigen herein.

An "immunological response" to an antigen or composition is the development in a subject of a humoral and/or a cellular immune response to an antigen present in the composition of interest. For purposes of the present invention, a "humoral immune response" refers to an immune response mediated by antibody molecules, including secretory (IgA) or IgG molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important aspect of cellular immunity involves an antigen-specific response by cytolytic T-cells ("CTL"s). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the surfaces of cells. CTLs help induce and promote the destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood cells, including those derived from CD4+ and CD8+ T-cells. In addition, a chemokine response may be induced by various white blood or endothelial cells in response to an administered antigen.

II. VACCINE FORMULATIONS

The invention relates to vaccine formulations for the treatment or prevention of Severe Acute Respiratory Syndrome (SARS). Vaccine formulations of the invention include an inactivated (or killed) SARS virus, an attenuated SARS virus, a split SARS virus preparation and a recombinant or purified subunit formulation of one or more SARS viral antigens. The invention includes polypeptides and polynucleotides encoding for SARS viral antigens and

immunogenic fragments thereof. Expression and delivery of the polynucleotides of the invention may be facilitated via viral vectors and/or viral particles, including Virus Like Particles (VLPs).

A. Inactivated (or Killed) SARS Vaccines

The invention includes a composition comprising an inactivated (or killed) SARS virus and methods for the production thereof. Inactivated SARS viral compositions can be used as prophylactic or therapeutic SARS virus vaccine. Preferably the inactivated SARS virus vaccine composition comprises an amount of inactivated SARS virus which, before inactivation, is equivalent to a virus titer of from about 4 to 7 logs plaque forming units (PFU) or 4 to 7 logs tissue culture infectious dose 50 ($TCID_{50}$) per milliliter. More preferably, before inactivation the virus titer is from 4 to 11, 7 to 11 or 9 to 11 PFU or $TCID_{50}$. Still more preferably the inactivated SARS virus vaccine composition comprises an amount of inactivated SARS virus which, before inactivation, is equivalent to a virus titer of from about 5 to 9 PFU or 5 to 9 $TCID_{50}$ per milliliter. In one embodiment, the PFU or $TCID_{50}$ of the cultured SARS virus at harvest is 6 to 8, more preferably about 7.5 PFU or $TCID_{50}$ per milliliter. Upon concentration of the viral harvest, the PFU or $TCID_{50}$ is preferably 8 to 11, still more preferably about 9 PFU or $TCID_{50}$ per milliliter. The vaccine composition comprises a sufficient amount of the SARS virus antigen to produce an immunological response in a primate.

Methods of inactivating or killing viruses are known in the art to destroy the ability of the viruses to infect mammalian cells. Such methods include both chemical or physical means. Chemical means for inactivating a SARS virus include treatment of the virus with an effective amount of one or more of the following agents: detergents, formaldehyde, formalin, β-propiolactone, or UV light. Additional chemical means for inactivation include treatment with methylene blue, psoralen, carboxyfullerene (C60) or a combination of any thereof. Other methods of viral inactivation are known in the art, such as for example binary ethylamine, acetyl ethyleneimine, or gamma irradiation.

For example formaldehyde may be used at concentrations such as 0.1 to 0.02%, preferably at 0.02 to 0.1 %, and still more preferably at 0.04 to 0.05%. The inactivating agent is added to virus containing culture supernatants prior to or after harvesting said culture supernatants from vessels used for virus propagation, either with or without a step of cell disruption for release of cell-associated virus prior to harvesting. Further, the inactivating agent may be added after said culture supernatants have been stored frozen and thawed, or after one or more steps of purification to remove cell contaminants. Preferably, however, formaldehyde is added after removal of cells and cellular debris or after one or more purification steps. After addition of formaldehyde, the virus containing mixture is transferred into an incubation vessel and incubated at refrigeration temperatures (e.g. +2 to 8°C) or alternatively at elevated temperatures, such as ambient temperatures between approximately 20 and 30°C or at 33°C to 37°C for a period of 12

hours to 7 days, whereby the temperature chosen should be adjusted to the duration of incubation. Prefered conditions are e.g. +2 –8°C for 3-7 days (prefered are 3 -7days), ambient temperatures and incubation for 16 hours to 3 days (prefered 24- 48 hours), or 35-37°C for 12-36 hours. If it is desirable to remove excess formalin, sodium thiosulfate or sodium metabisulfite at equimolar or 1.5 -fold molar concentration (relative to formaldehyde) may be added after completing the inactivation process.

For example, β-propiolactone may be used at concentrations such as 0.01 to 0.5%, preferably at 0.5% to 0.2%, and still more preferably at 0.025 to 0.1%. The inactivating agent is added to virus containing culture supernatants (virus material) prior to or after harvesting said culture supernatants from vessels used for virus propagation, either with or without a step of cell disruption for release of cell-associated virus prior to harvesting. Further, the inactivating agent may be added after said culture supernatants have been stored frozen and thawed, or after one or more steps of purification to remove cell contaminants. β-propiolactone is added to the virus material, with the adverse shift in pH to acidity being controlled with sodium hydroxide (*e.g.*, 1 N NaOH), a Tris-buffer or sodium bicarbonate solution. After transfering the mixture to another inactivation vessel, the combined inactivating agent-virus materials are incubated at temperatures from 4°C to 37°C, for incubation times of preferably 24 to 72 hours.

Another inactivant which may be used is binary ethyleneimine (BEI). Equal volumes of a 0.2 molar bromoethylamine hydrobromide solution and a 0.4 molar sodium hydroxide solution are mixed and incubated at about 37°C. for 60 minutes. The resulting cyclized inactivant is binary ethyleneimine, which is added to the virus materials at 0.5 to 4 percent, and preferably at 1 to 3 percent, volume to volume. The inactivating virus materials are held from about 4°C to 37°C for 24 to 72 hours with periodic agitation. At the end of this incubation 20 ml. of a sterile 1 molar sodium thiosulfate solution was added to insure neutralization of the BEI.

In one embodiment, the invention includes an inactivating method is designed to maximize exposure of the virus to the inactivating agent and to minimize long-term exposure of the temperature sensitive SARS virus particles to elevated temperatures. The invention includes an inactivation method comprising exposing the virus to the inactivation agent (such as BPL) for 12 to 24 hours at refrigeration temperatures followed by hydrolysis of any residual inactivating agent by elevating the temperature for only 3 hours. Preferably, the refrigeration temperatures are between 0 and 8°C, more preferably around 4°C. Preferably, the elevated temperature is between 33 and 41°C, more preferably around 37°C. As assessed by a test for residual infectious virus using 10 ml aliquots of the inactivated preparation, the method is able to inactivate SARS-CoV in raw cell culture harvests below a theoretical limit of 0.03 infectious units/ml.

Diluted and undiluted samples of the inactivated virus materials are added to susceptible cell (tissue) culture (*e.g.*, VERO) to detect any non-inactivated virus. The cultured cells are

passaged multiple times and examined for the presence of SARS virus based on any of a variety of methods, such as, for example, cytopathic effect (CPE) and antigen detection (*e.g.*, via fluoroscent antibody conjugates specific for SARS virus). Such tests allow determination of complete virus inactivation.

Prior to inactivation, the SARS virus will be cultured in a mammalian cell culture. The cell culture may be adherently growing cells or cells growing in suspension. Preferably the cells are of mammalian origin, but may also be derived from avian (*e.g.*, hens' cells such as hens' embryo cells (CEF cells)), amphibian, reptile, insect, or fish sources. Mammalian sources of cells include, but are not limited to, human or non-human primate (*e.g.*, MRC-5 (ATCC CCL-171), WI-38 (ATCC CCL-75), HeLa cells, human diploid cells, fetal rhesus lung cells (*e.g.* ATCC CL-160), human embryonic kidney cells (293 cells, typically transformed by sheared adenovirus type 5 DNA), VERO cells (*e.g.*, from monkey kidneys), horse, cow (*e.g.*, MDBK cells), sheep, dog (*e.g.*, MDCK cells from dog kidneys, ATCC CCL34 MDCK (NBL2) or MDCK 33016, deposit number DSM ACC 2219 as described in WO 97/37001), cat, and rodent (*e.g.*, hamster cells such as BHK21-F, HKCC cells, or Chinese hamster ovary cells (CHO cells)), and may be obtained from a wide variety of developmental stages, including for example, adult, neonatal, fetal, and embryo.

In certain embodiments the cells are immortalized (*e.g.*, PERC.6 cells are described, for example, in WO 01/38362 and WO 02/40665, incorporated by reference herein in their entireties, as well as deposited under ECACC deposit number 96022940), or any other cell type immortalized using the techniques described herein.

In preferred embodiments, mammalian cells are utilized, and may be selected from and/or derived from one or more of the following non-limiting cell types: fibroblast cells (*e.g.*, dermal, lung), endothelial cells (*e.g.*, aortic, coronary, pulmonary, vascular, dermal microvascular, umbilical), hepatocytes, keratinocytes, immune cells (*e.g.*, T cell, B cell, macrophage, NK, dendritic), mammary cells (*e.g.*, epithelial), smooth muscle cells (*e.g.*, vascular, aortic, coronary, arterial, uterine, bronchial, cervical, retinal pericytes), melanocytes, neural cells (*e.g.*, astrocytes), prostate cells (*e.g.*, epithelial, smooth muscle), renal cells (*e.g.*, epithelial, mesangial, proximal tubule), skeletal cells (*e.g.*, chondrocyte, osteoclast, osteoblast), muscle cells (*e.g.*, myoblast, skeletal, smooth, bronchial), liver cells, retinoblasts, and stromal cells. WO 97/37000 and WO 97/37001, incorporated by reference herein in their entireties, describe production of animal cells and cell lines that capable of growth in suspension and in serum free media and are useful in the production and replication of viruses.

Preferably, the SARS viruses of the invention are grown on VERO cells or fetal rhesus kidney cells.

Culture conditions for the above cell types are well-described in a variety of publications, or alternatively culture medium, supplements, and conditions may be purchased commercially, such as for example, as described in the catalog and additional literature of Cambrex Bioproducts (East Rutherford, NJ).

In certain embodiments, the host cells used in the methods described herein are cultured in serum free and/or protein free media. A medium is referred to as a serum-free medium in the context of the present invention in which there are no additives from serum of human or animal origin. Protein-free is understood to mean cultures in which multiplication of the cells occurs with exclusion of proteins, growth factors, other protein additives and non-serum proteins. The cells growing in such cultures naturally contain proteins themselves.

Known serum-free media include Iscove's medium, Ultra-CHO medium (BioWhittaker) or EX-CELL (JRH Bioscience). Ordinary serum-containing media include Eagle's Basal Medium (BME) or Minimum Essential Medium (MEM) (Eagle, Science, 130, 432 (1959)) or Dulbecco's Modified Eagle Medium (DMEM or EDM), which are ordinarily used with up to 10% fetal calf serum or similar additives. Optionally, Minimum Essential Medium (MEM) (Eagle, Science, 130, 432 (1959)) or Dulbecco's Modified Eagle Medium (DMEM or EDM) may be used without any serum containing supplement. Protein-free media like PF-CHO (JHR Bioscience), chemically-defined media like ProCHO 4CDM (BioWhittaker) or SMIF 7 (Gibco/BRL Life Technologies) and mitogenic peptides like Primactone, Pepticase or HyPep™ (all from Quest International) or lactalbumin hydrolyzate (Gibco and other manufacturers) are also adequately known in the prior art. The media additives based on plant hydrolyzates have the special advantage that contamination with viruses, mycoplasma or unknown infectious agents can be ruled out.

The cell culture conditions to be used for the desired application (temperature, cell density, pH value, *etc.*) are variable over a very wide range owing to the suitability of the cell line employed according to the invention and can be adapted to the requirements of the SARS virus.

The method for propagating the SARS virus in cultured cells (*e.g.*, mammalian cells) includes the steps of inoculating the cultured cells with SARS virus, cultivating the infected cells for a desired time period for virus propagation, such as for example as determined by SARS virus titer or SARS virus antigen expression (*e.g.*, between 24 and 168 hours after inoculation) and collecting the propagated virus. The cultured cells are inoculated with a SARS virus (measured by PFU or $TCID_{50}$) to cell ratio of 1:10000 to 1:10. A lower range of ratios may also be used *e.g.* 1:500 to 1:1, preferably 1:100 to 1:5, more preferably 1:50 to 1:10. The SARS virus is added to a suspension of the cells or is applied to a monolayer of the cells, and the virus is absorbed on the cells for at least 60 minutes but usually less than 300 minutes, preferably between 90 and 240 minutes at 25°C to 40°C, more preferably 28°C to 37°C, still more

preferably at about 33 °C. The infected cell culture (*e.g.*, monolayers) may be treated either by freeze-thawing or by enzymatic action to increase the viral content of the harvested culture supernatants. The harvested fluids are then either inactivated or stored frozen.

A comparison of SARS infected Vero cells grown with and without fetal calf serum ("FCS") is shown in FIGURE 26A. Briefly, Vero cells were split the day before infection and cultivated in T175 flasks. Infection of 90% confluent Vero cell monolayers the following day was performed with a SARS-CoV seed stock (strain FRA, passage 4, Accession number AY310120), with or without 3% FCS (Fig. 26A). The addition of FCS to the cell media showed little impact on virus yield.

Cultured cells may be infected at a multiplicity of infection ("m.o.i.") of about 0.0001 to 10, preferably 0.002 to 5, more preferably to 0.001 to 2. Still more preferably, the cells are infected at a m.o.i of about 0.01. A comparison of viral yield at varying m.o.i. levels is shown in FIGURE 26B.

Infected cells may be harvested 30 to 60 hours post infection. Preferably, the cells are harvested 34-48 hours post infection. Still more preferably, the cells are harvested 38 to 40 hours post infection. See FIGURE 26C.

Methods of purification of inactivated virus are known in the art and may include one or more of, for instance gradient centrifugation, ultracentrifugation, continuous-flow ultracentrifugation and chromatography, such as ion exchange chromatography, size exclusion chromatography, and liquid affinity chromatography. Additional method of purification include ultrafiltration and dialfiltration. *See* JP Gregersen "Herstellung von Virussimpfstoffen aus Zellkulturen" Chapter 4.2 in Pharmazeutische Biotecnologie (eds. O. Kayser and RH Mueller) Wissenschaftliche Verlagsgesellschaft, Stuttgart, 2000. *See also*, O'Neil *et al.*, "Virus Harvesting and Affinity Based Liquid Chromatography. A Method for Virus Concentration and Purification", Biotechnology (1993) 11:173-177; Prior *et al.*, "Process Development for Manufacture of Inactivated HIV-1", Pharmaceutical Technology (1995) 30-52; and Majhdi *et al.*, "Isolation and Characterization of a Coronavirus from Elk Calves with diarrhea" Journal of Clinical Microbiology (1995) 35(11): 2937-2942.

Other examples of purification methods suitable for use in the invention include polyethylene glycol or ammonium sulface precipitation (*see* Trepanier *et al.*, "Concentration of human respiratory syncytial virus using ammonium sulfate, polyethylene glycol or hollow fiber ultrafiltration" Journal of Virological Methods (1981) 3(4):201-211; Hagen *et al.*, "Optimization of Poly(ethylene glycol) Precipitation of Hepatitis Virus Used to prepare VAQTA, a Highly Purified Inactivated Vaccine" Biotechnology Progress (1996) 12:406-412; and Carlsson *et al.*, "Purification of Infectious Pancreatic Necrosis Virus by Anion Exchange Chromatography Increases the Specific Infectivity" Journal of Virological Methods (1994) 47:27-36) as well as

ultrafiltration and microfiltration (*see* Pay *et al.*, Developments in Biological Standardization (1985) 60:171-174; Tsurumi *et al.*, "Structure and filtration performances of improved cuprammonium regenerated cellulose hollow fibre (improved BMM hollow fibre) for virus removal" Polymer Journal (1990) 22(12):1085-1100; and Makino *et al.*, "Concentration of live retrovirus with a regenerated cellulose hollow fibre, BMM", Archives of Virology (1994) 139(1-2):87-96.).

Preferably, the virus is purified using chromatography, such as ion exchange chromatography. Chromatic purification allows for the production of large volumes of virus containing suspension. The viral product of interest can interact with the chromatic medium by a simple adsorption/desorption mechanism, and large volumes of sample can be processed in a single load. Contaminants which do not have affinity for the adsorbent pass through the column. The virus material can then be eluted in concentrated form.

Preferred anion exchange resins for use in the invention include DEAE, EMD TMAE. Preferred cation exchange resins may comprise a sulfonic acid-modified surface. In one embodiment, the virus is purified using ion exchange chromatography comprising a strong anion exchange resin (*e.g.* EMD TMAE) for the first step and EMD-SO_3 (cation exchange resin) for the second step. A metal-binding affinity chromatography step can optionally be included for further purification. (See, *e.g.*, WO 97/06243).

A preferred resin for use in the invention is Fractogel™ EMD. This synthetic methacrylate based resin has long, linear polymer chains (so-called "tentacles") covalently attached. This "tentacle chemistry" allows for a large amount of sterically accessible ligands for the binding of biomolecules without any steric hindrance. This resin also has improved pressure stability.

Column-based liquid affinity chromatography is another preferred purification method for use in the invention. One example of a resin for use in this purification method is Matrex™ Cellufine™ Sulfate (MCS). MCS consists of a rigid spherical (approx. 45-105 μm diameter) cellulose matrix of 3,000 Dalton exclusion limit (its pore structure excludes macromolecules), with a low concentration of sulfate ester functionality on the 6-position of cellulose. As the functional ligand (sulfate ester) is relatively highly dispersed, it presents insufficient cationic charge density to allow for most soluble proteins to adsorb onto the bead surface. Therefore the bulk of the protein found in typical virus pools (cell culture supernatants, *e.g.* pyrogens and most contaminating proteins, as well as nucleic acids and endotoxins) are washed from the column and a degree of purification of the bound virus is achieved.

The rigid, high-strength beads of MCS tend to resist compression. The pressure/flow characteristics the MCS resin permit high linear flow rates allowing high-speed processing, even in large columns, making it an easily scalable unit operation. In addition a chromatographic purification step with MCS provides increased assurance of safety and product sterility, avoiding

excessive product handling and safety concerns. As endotoxins do not bind to it, the MCS purification step allows a rapid and contaminant free depyrogenation. Gentle binding and elution conditions provide high capacity and product yield. The MCS resin therefore represents a simple, rapid, effective, and cost-saving means for concentration, purification and depyrogenation. In addition, MCS resins can be reused repeatedly.

The inactivated virus may be further purified by gradient centrifugation, preferably density gradient centrifugation. For commercial scale operation a continuous flow sucrose gradient centrifugation would be the preferred option. This method is widely used to purify antiviral vaccines and is known to the expert in the field (*See* JP Gregersen "Herstellung von Virussimpfstoffen aus Zellkulturen" Chapter 4.2 in Pharmazeutische Biotecnologie (eds. O. Kayser and RH Mueller) Wissenschaftliche Verlagsgesellschaft, Stuttgart, 2000.)

The density gradient centrifugation step may be performed using laboratory or commercial scale gradient centrifugation equipment. For example, a swinging bucket rotor, a fixed angle rotor, or a vertical tube rotor, particularly for laboratory scale production of the virus. Preferably, the gradient centrifugation step is performed using a swinging bucket rotor. This type of rotor has a sufficiently long pathlength to provide high quality separations, particularly with multicomponent samples. In addition, swinging bucket rotors have greatly reduced wall effects, and the contents do not reorient during acceleration and deceleration. Because of their longer pathlength, separations take longer compared to fixed angle or vertical tube rotors. The prepared sucrose solutions are controlled via refractometer on their sucrose concentration.

Sucrose gradients for density gradient centrifugation, such as in a swinging bucket centrifuge tubes may be formed prior to centrifugation by the use of a gradient former (continuous/linear). The volume of sample which can be applied to the gradient in a swinging bucket rotor tube is a function of the cross-sectional area of the gradient that is exposed to the sample. If the sample volume is too high, there is not sufficient radial distance in the centrifuge tube for effective separation of components in a multicomponent sample.

An approximate sample volume for swinging bucket rotor SW 28 is 1-5 ml per tube (with a tube diameter of 2.54 cm). The sample is applied to the gradient by pipetting the volume on top of the gradient. The blunt end of the pipette is placed at 45-60° angle to the tube wall, approximately 2-3 mm above the gradient. The sample is injected slowly and allowed to run down the wall of the tube onto the gradient. After centrifugation gradient fractions are recovered by carefully inserting a gauge needle until the bottom of the tube and starting to collect fractions of 2 ml by pumping the liquid from the tube into falcon tubes.

Sucrose density gradients suitable for use with this density gradient centrifugation purification step include 0 – 60%, 5 – 60%, 15 – 60%, 0 – 50%, 5 – 50%, 15 – 50%, 0 – 40%, 5 – 40%, and 15 – 40%. Preferably, the sucrose density gradient is 15 – 40%, 5 – 40% or 0 – 40%.

Alternatively, a discontinuous sucrose density gradient may be used for purification. A discontinuous sucrose density scheme provides for discrete, overlaying layers of differing sucrose concentrations. In one example, a first layer of 50% sucrose is covered by a second layer of 40% sucrose; the second layer is covered by a third layer of 20% sucrose; the third layer is covered by a fourth layer of 10% sucrose; and the fourth layer is covered by the solution containing the virus to be purified.

In one embodiment, inactivated virus is purified by a method comprising a first step of chromatography purification and a second step of gradient centrifugation. Preferably the first step comprises liquid affinity chromatography, such as MCS. Preferably, the second step comprises density gradient centrifugation using a swinging bucket rotor.

Additional purification methods which may be used to purify inactivated SARS virus include the use of a nucleic acid degrading agent, preferably a nucleic acid degrading enzyme, such as a nuclease having DNase and RNase activity, or an endonuclease, such as from *Serratia marcescens*, commercially available as Benzonase™, membrane adsorbers with anionic functional groups (*e.g.* Sartobind™) or additional chromatographic steps with anionic functional groups (*e.g.* DEAE or TMAE). An ultrafiltration/dialfiltration and final sterile filtration step could also be added to the purification method.

Preferably, the purification includes treatment of the SARS viral isolate with one or more nucleic acid degrading enzymes. These enzymes may be used to reduce the level of host cell nucleic acid in the viral purification process. Nucleic acid digesting enzymes for use in cell culture are known in the art and include, for example, Benzonase™.

The treatment of the virus with the nucleic acid degrading enzyme and inactivating agent can be performed by a sequential treatment or in a combined or simultaneous manner. Preferably, the nucleic acid degrading agent is added to the virus preparation prior to the addition of the inactivating agent.

The purified viral preparation of the invention is substantially free of contaminating proteins derived from the cells or cell culture and preferably comprises less than about 1000, 500, 250, 150, 100, or 50 pg cellular nucleic acid / μg virus antigen, preferably less than about 1000, 500, 250, 150, 100, or 50 pg cellular nucleic acid/ dose. Still more preferably, the purified viral preparation comprises less than about 20 pg, and even more preferably, less than about 10 pg. Methods of measuring host cell nucleic acid levels in a viral sample are known in the art. Standardized methods approved or recommended by regulatory authorities such as the WHO or the FDA are preferred.

The invention includes an inactivated vaccine composition comprising a prophylactically effective amount of SARS viral antigen, preferably spike or an immunogenic fragment thereof. The SARS viral antigen is preferably present in a concentration amount of 0.1 to 50 μg

antigen/dose, more preferably 0.3 to 30 μg antigen/dose. Still more preferably, the antigen is about 15 μg/dose.

In one embodiment, a lower concentration of SARS viral antigen is used in inactivated vaccine compositions of the invention. Such lower concentration vaccines may optionally comprise an adjuvant to boost the host immune response to the antigen. In such a "low dose" vaccine, the SARS viral antigen is preferably present in a concentration of less than 15 μg antigen/dose, (*i.e.*, less than 10, 7.5, 5 or 3 μg antigen/dose.

The inactivated vaccine preparations of the invention may further comprise a stabilizer to preserve the integrity of the immunogenic proteins in the inactivated viral preparation. Stabilizers suitable for use in vaccines are known in the art and may include, for example, buffers, sugars, sugar alcohols, and amino acids. Stabilizing buffers are preferably adjusted to a physiological pH range and may include phosphate buffers, Tris buffers, TE (Tris/EDTA), TEN (Tris/NaCl/EDTA) and Earle's salt solution. Stabilizing sugars may include, for example, one or more of saccharose, glucose, fructose, dextranes, dextranesulphate, and trehalose. Stabilizing sugar alcohols may include, for example, Xylite/Xylitole, Mannite/Mannitol, Sorbite/Sorbitol, and Glycerol. Amino acids suitable for use in the invention include, for example, L-glutamine, arginine, cysteine, and lysine. Additional stabilizers which may be used in the invention include Tartaric acid, Pluronic F 68, and Tween 80.

SARS viral isolates which may be used for the inactivated viral preparations of the invention may be obtained and identified by any of the mechanisms described supra. For example, a SARS isolate may be obtained from a clinical sample and plaque purified. Such methods of viral isolation are known in the art.

Further purification procedures can be applied to ensure the seed virus used for preparation of the vaccine does not contain, for example, unwanted adventitious agents. In one embodiment, viral RNA from the viral isolate can be isolated from the virus, purified (and, optionally, the sequence verified through PCR or other means) and then introduced into a suitable cell culture.

As an example of this technique, a clinical viral sample is plaque purified and amplified on vero cells to generate a sufficient amount of the viral sample for analysis. Cellular remnants are then cleared from the supernatant by centrifugation. The virus can then be pelleted by ultracentrifugation and the pellet resuspended in PBS. After further centrifugation purification, the virus containing fraction is treated with a DNase (and optionally also an RNase). Viral RNA is then isolated from this fraction and transfected into a host cell.

Examples 2 and 3 provide an illustration of purification of inactivated whole SARS virus using MCS chromatography resin purification followed by density gradient ultracentrifugation.

Routes and methods of immunization of the vaccines of the invention are discussed in more detail in a section below. Examples 4 and 5 provide illustrations of a mouse immunization scheme with the inactivated SARS virus of the invention.

B. Attenuated SARS Vaccines

The invention includes a composition comprising an attenuated SARS virus. This composition can be used as a prophylactic or therapeutic SARS virus vaccine. Methods of attenuating viruses are known in the art. Such methods include serial passage of the SARS virus in cultured cells (*e.g.*, mammalian cell culture, preferably fetal rhesus kidney cells or VERO cells-see the discussion in Section A above regarding culture of SARS virus), until the SARS virus demonstrates attenuated function. The temperature at which the virus is grown can be any temperature at which with tissue culture passage attenuation occurs. Attenuated function of the SARS virus after one or more passages in cell culture can be measured by one skilled in the art. As used herein, attenuation refers to the decreased virulence of the SARS virus in a human subject. Evidence of attenuated function may be indicated by decreased levels of viral replication or by decreased virulence in an animal model.

Other methods of producing an attenuated SARS virus include passage of the virus in cell culture at sub-optimal or "cold" temperatures and introduction of attenuating mutations into the SARS viral genome by random mutagenesis (*e.g.*, chemical mutagenesis) or site specific directed mutagenesis. Preparation and generation of attenuated RSV vaccines (the methods of which will generally applicable to SARS virus) are disclosed in, for example, EP 0 640 128, US Patent No. 6,284,254, US Patent No. 5,922,326, US Patent No. 5,882,651.

The attenuated derivatives of SARS virus are produced in several ways, such as for example, by introduction of temperature sensitive-mutations either with or without chemical mutagenesis (*e.g.*, 5-fluorouracil), by passage in culture at "cold" temperatures. Such cold adaptation includes passage at temperatures between about 20°C to about 32°C, and preferably between temperatures of about 22°C to about 30°C, and most preferably between temperatures of about 24°C and 28°C. The cold adaptation or attenuation may be performed by passage at increasingly reduced temperatures to introduce additional growth restriction mutations. The number of passages required to obtain safe, immunizing attenuated virus is dependent at least in part on the conditions employed. Periodic testing of the SARS virus culture for virulence and immunizing ability in animals (*e.g.*, mouse, primate) can readily determine the parameters for a particular combination of tissue culture and temperature. The attenuated vaccine will typically be formulated in a dose of from about 10^3 to 10^6 PFU or $TCID_{50}$, or more for maximal efficacy.

Attenuated virus vaccines for SARS-CoV also are produced by creating virus chimeras comprising sequences derived from at least two different coronaviruses, one of which is a SARS-CoV. For example, a virus chimera is produced that comprises nonstructural protein encoding

genes derived from a first coronavirus (e.g., murine, bovine, porcine, canine, feline, avian coronavirus) and one or more structural protein encoding genes (*e.g.*, spike, E, M) from a SARS-CoV. Alternatively, the virus chimera may comprise sequences derived from a human coronavirus that is not a SARS-CoV (*e.g.*, OC43, 229E) together with sequences from a SARS-CoV. Chimeric coronaviruses of the present invention are generated by a variety of methods, including for example allowing for natural RNA recombination in a eukaryotic (*e.g.*, mammalian) cell that contains RNA from each of the parental coronaviruses (*e.g.*, following infection) or by using standard molecular biology techniques known to those of skill in the art to engineer desired virus chimeras (or portions thereof) as cDNA clones, which may then be used to produce infectious virus (see for example, US 6593111 B2; Yount *et al.*, 2003, *Proc. Natl. Acad. Sci. USA* 100(22):12995-13000). An attenuated phenotype of the coronavirus chimeras described herein can be readily measured by one of skill in the art.

Attenuated viruses can be also generated by deleting one or more open reading frames (ORFs) that are not essential for viral replication. Preferably, these deletions occur in the structural region of the genome, such as ORF 3a, 3b, 6, 7a, 7b, 8a, 8b, 9b. See e.g., Haijema BJ, Volders H, Rottier PJ. J Virol. (2004) 78(8):3863-71; and de Haan, C. A., P. S. Masters, X. Shen, S. Weiss, and P. J. Rottier, "The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host." Virology (2002) 296:177-189. Deletion of such regions within a coronavirus such as SARS can be achieved, for example, by reverse genetics or "targeted recombination" (See, e.g., Masters, P. S., "Reverse genetics of the largest RNA viruses", Adv. Virus Res. (1999) 53:245-264.

Methods of purification of attenuated virus are known in the art and may include one or more of, for instance gradient centrifugation and chromatography. See Gregersen "Herstellung von Virussimpfstoffen aus Zellkulturen" Chapter 4.2 in Pharmazeutische Biotecnologie (eds. O. Kayser and RH Mueller) Wissenschaftliche Verlagsgesellschaft, Stuttgart, 2000.

C. Split SARS Vaccines

The invention includes a composition comprising a split SARS virus formulation and methods for the manufacture thereof. This composition can be used as a prophylactic or therapeutic SARS virus vaccine.

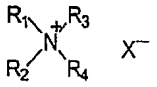
Methods of splitting enveloped viruses are known in the art. Methods of splitting enveloped viruses are disclosed, for example, in WO 02/28422, incorporated herein by reference in its entirety, and specifically including the splitting agents and methods described therein. Methods of splitting influenza viruses are disclosed, for example, in WO 02/067983, WO 02/074336, and WO 01/21151, each of which is incorporated herein by reference in its entirety.

The splitting of the virus is carried out by disrupting or fragmenting whole virus, infectious (wild-type or attenuated) or non-infectious (for example inactivated), with a disrupting

concentration of a splitting agent. The disruption results in a full or partial solubilisation of the virus proteins, altering the integrity of the virus.

Preferably, the splitting agent is a non-ionic or an ionic surfactant. Accordingly, the split SARS virus formulations of the invention may also comprise at least one non-ionic surfactant or detergent. Examples of splitting agents useful in the invention include: bile acids and derivatives thereof, non-ionic surfactants, alkylglycosides or alkylthioglycosides and derivatives thereof, acyl sugars, sulphobetaines, betains, polyoxyethylenealkylethers, N,N-dialkyl-Glucamides, Hecameg, alkylphenoxypolyethoxyethanols, quaternary ammonium compounds, sarcosyl, CTAB (cetyl trimethyl ammonium bromide) or Cetavlon.

Preferably, the ionic surfactant is a cationic detergent. Cationic detergents suitable for use in the invention include detergents comprising a compound of the following formula:



wherein

R_1, R_2 and R_3 are the same or different and each signifies alkyl or aryl, or

R_1 and R_2, together with the nitrogen atom to which these are attached form a 5- or 6-membered heterocyclic ring, and

R_3 signifies alkyl or aryl, or

R_1, R_2 and R_3 together with the nitrogen atom to which these are attached, signify a 5- or 6-membered heterocyclic ring, unsaturated at the nitrogen atom,

R_4 signfies alkyl or aryl, and

X signifies an anion.

Examples of such cationic detergents are cetyltrimethylammonium salts, such as ceytltrimethylammonium bromide (CTAB) and myristyltrimethylammonium salt.

Additional cationic detergents suitable for use in the invention include lipofectine, lipofectamine, and DOT-MA.

Non-ionic surfactants suitable for use in the invention include one or more selected from the group consisting of the octyl- or nonylphenoxy polyoxyethanols (for example the commercially available Triton series), polyoxyethylene sorbitan esters (Tween series) and polyoxyethylene ethers or esters of the general formula :

$$O(CH_2CH_2O)_n\text{-}A\text{-}R$$

wherein n is 1-50, A is a bond or –C(O)-, R is $C_{1\text{-}50}$ alkyl or phenyl $C_{1\text{-}50}$ alkyl; and combinations of two or more of these.

The invention comprises a method of preparing a split SARS virus comprising contacting the SARS virus with a sufficient amount of splitting agent to disrupt the viral envelope. The loss

of integrity after splitting renders the virus non-infectious. Once the disrupted viral envelope proteins are generally no longer associated with whole intact virions, other viral proteins are preferably fully or partially solubilized and are therefore not associated, or only in part associated, with whole intact virions after splitting.

The method of preparing a split SARS virus may further comprise removal of the splitting agents and some or most of the viral lipid material. The process may also include a number of different filtration and/or other separation steps such as ultracentrifugation, ultrafiltration, zonal centrifugation and chromatographic steps in a variety of combinations. The process may also optionally include an inactivation step (as described above) which may be carried out before or after the splitting. The splitting process may be carried out as a batch, continuous, or semi-continuous process.

Split SARS virus vaccines of the invention may include structual proteins, membrane fragments and membrane envelope proteins. Preferably, the split SARS virus preparations of the invention comprise at least half of the viral structural proteins.

One example of a method of preparing a split SARS virus formulation includes the following steps:

(i) propagation of the SARS virus in cell culture, such as MRC-5 cells (ATCC CCL-171), WI-38 cells (ATCC CCL-75), fetal rhesus kidney cells or vero cells (See the discussion in Section A, above, regarding culture of SARS virus);

(ii) harvesting of SARS virus-containing material from the cell culture;

(iii) clarification of the harvested material to remove non-SARS virus material;

(iv) concentration of the harvested SARS virus;

(v) separation of the whole SARS virus from non-virus material;

(vi) splitting of the whole SARS virus using a suitable splitting agent in a density gradient centrifugation step; and

(vii) filtration to remove undesired materials.

The above steps are preferably performed sequentially.

The clarification step is preferably performed by centrifugation at a moderate speed. Alternatively, a filtration step may be used for example with a 0.2μm membrane.

The concentration step may preferably employ an adsorption method, for instance, using $CaHPO_4$. Alternatively, filtration may be used, for example ultrafiltration.

A further separation step may also be used in the method of the invention. This further separation step is preferably a zonal centrifugation separation, and may optionally use a sucrose gradient. The sucrose gradient may further comprise a preservative to prevent microbial growth.

The splitting step may also be performed in a sucrose gradient, wherein the sucrose gradient contains the splitting agent.

The method may further comprise a sterile filtration step, optionally at the end of the process. Preferably, there is an inactivation step prior to the final filtration step.

Methods of preparing split SARS virus formulations may further include treatment of the viral formulation with a DNA digesting enzyme. These enzymes may be used to reduce the level of host cell DNA in the viral purification process. DNA digesting enzymes for use in cell culture are known in the art and include, for example, Benzonase®.

Treatment of the SARS virus formulation with a DNA digesting enzyme may occur at any time in the purification and splitting process. Preferably, however, the SARS virus formulation is treated with a DNA digesting enzyme prior to use of a detergent. Still more preferably, the SARS virus formulation is treated with a DNA digesting enzyme, such as Benzonas, prior to treatment with a cationic detergent, such as CTAB.

Methods of purification of split virus are known in the art. See JP Gregersen "Herstellung von Virussimpfstoffen aus Zellkulturen" Chapter 4.2 in Pharmazeutische Biotecnologie (eds. O. Kayser and RH Mueller) Wissenschaftliche Verlagsgesellschaft, Stuttgart, 2000.

The invention includes a split vaccine composition comprising a prophylactically effective amount of SARS viral antigen, preferably spike or an immunogenic fragment thereof. The SARS viral antigen is preferably present in a concentration amount of 0.1 to 50 μg antigen/dose, more preferably 0.3 to 30 μg antigen/dose. Still more preferably, the antigen is about 15 μg/dose.

In one embodiment, a lower concentration of SARS viral antigen is used in split vaccine compositions of the invention. Such lower concentration vaccines may optionally comprise an adjuvant to boost the host immune response to the antigen. In such a "low dose" vaccine, the SARS viral antigen is preferably present in a concentration of less than 15 μg antigen/dose, (*i.e.*, less than 10, 7.5, 5 or 3 μg antigen/dose.

D. Subunit SARS Vaccines

The invention includes a composition comprising an isolated or purified SARS viral antigen or a derivative thereof. The composition may further comprise one or more adjuvants.

SARS viral antigens can be isolated or purified from a SARS virus grown in cell culture. Alternatively, SARS viral antigens can be recombinantly produced by methods known in the art.

The SARS viral antigens used in the invention can be produced in a variety of different expression systems which are known in the art; for example those used with mammalian cells, baculoviruses, bacteria, and yeast. Such expression systems will typically use polynucleotides encoding the viral antigens of the invention. Such sequences can be obtained using standard techniques of molecular biology, including translating the amino acid sequences listed herein. Accordingly, the invention includes polynucleotides encoding for the viral antigens of the

invention. In addition, the viral antigens of the invention can be produced (at least in part, preferably in whole) via synthetic chemistry methods.

Insect cell expression systems, such as baculovirus systems, are known to those of skill in the art and described in, *e.g.*, Summers and Smith, *Texas Agricultural Experiment Station Bulletin* No. 1555 (1987). Materials and methods for baculovirus/insert cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA. Similarly, bacterial and mammalian cell expression systems are also known in the art and described in, *e.g.*, *Yeast Genetic Engineering* (Barr *et al.*, eds., 1989) Butterworths, London.

A number of appropriate host cells for use with the above systems are also known. For example, mammalian cell lines are known in the art and include immortalized cell lines available from the American Type Culture Collection (ATCC), such as, but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (*e.g.*, Hep G2), Madin-Darby bovine kidney ("MDBK") cells, as well as others. Mammalian sources of cells include, but are not limited to, human or non-human primate (*e.g.*, MRC-5 (ATCC CCL-171), WI-38 (ATCC CCL-75), fetal rhesus lung cells (ATCC CL-160), human embryonic kidney cells (293 cells, typically transformed by sheared adenovirus type 5 DNA), VERO cells from monkey kidneys), horse, cow (*e.g.*, MDBK cells), sheep, dog (*e.g.*, MDCK cells from dog kidneys, ATCC CCL34 MDCK (NBL2) or MDCK 33016, deposit number DSM ACC 2219 as described in WO 97/37001), cat, and rodent (*e.g.*, hamster cells such as BHK21-F, HKCC cells, or Chinese hamster ovary cells (CHO cells)), and may be obtained from a wide variety of developmental stages, including for example, adult, neonatal, fetal, and embryo.

Similarly, bacterial hosts such as *E. coli*, *Bacillus subtilis*, and *Streptococcus spp.*, will find use with the present expression constructs. Yeast hosts useful in the present invention include, *inter alia*, *Saccharomyces cerevisiae*, *Candida albicans*, *Candida maltosa*, *Hansenual polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guillerimondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Insect cells for use with baculovirus expression vectors include, *inter alia*, *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni*.

Nucleic acid molecules comprising nucleotide sequences of the viral antigens or antibodies of the invention can be stably integrated into a host cell genome or maintained on a stable episomal element in a suitable host cell using various gene delivery techniques well known in the art. See., *e.g.*, US Patent No. 5,399,346.

Depending on the expression system and host selected, the molecules are produced by growing host cells transformed by an expression vector under conditions whereby the protein is expressed. The expressed protein is then isolated from the host cells and purified. If the expression system secretes the protein into growth media, the product can be purified directly

from the media. If it is not secreted, it can be isolated from cell lysates. The selection of the appropriate growth conditions and recovery methods are within the skill of the art.

The invention includes a composition comprising an isolated or purified SARS viral antigen or a derivative thereof. The invention also includes a composition comprising at least two isolated or purified SARS viral antigens or derivatives thereof, which have been co-purified or purified separately and then combined. In one embodiment, the SARS viral antigen is a spike (S) protein. In yet another embodiment, the SARS viral antigen is a nucleocapsid (N) protein, a membrane (M) glycoprotein, or an envelope (E) protein. Preferably, the SARS viral antigen is present in the composition in a purity greater than 75% (*e.g.*, 78%, 80%, 82%, 85%, 88%, 90%, 92%, 95%, 98%).

The invention includes a vaccine composition comprising a prophylactically effective amount of SARS viral antigen, preferably spike or an immunogenic fragment thereof. The SARS viral antigen is preferably present in a concentration amount of 0.1 to 50 μg antigen/dose, more preferably 0.3 to 30 μg antigen/dose. Still more preferably, the antigen is about 15 μg/dose.

In one embodiment, a lower concentration of SARS viral antigen is used in vaccine compositions of the invention. Such lower concentration vaccines may optionally comprise an adjuvant to boost the host immune response to the antigen. In such a "low dose" vaccine, the SARS viral antigen is preferably present in a concentration of less than 15 μg antigen/dose, (*i.e.*, less than 10, 7.5, 5 or 3 μg antigen/dose.

The following example illustrates a method of preparing a SARS virus spike (S) protein subunit vaccine.

SARS virus S antigen may be isolated and purified from a variety of sources and using a variety of methods, including, but not limited to, S antigen expressed in cultured eukaryotic cells (*e.g.*, mammalian cells, such as VERO, CHO) or bacteria (*e.g.*, *E. coli*). Expression of may be achieved by a variety of means, such as, for example, from SARS virus infected cell culture or cell culture supernatants, from cultured cells stably transformed with a DNA expression cassette encoding the SARS virus S protein (*e.g.*, RNA polymerase II promoter operably linked to a SARS virus S gene), or from cultured cells infected with a replication-competent or replication-incompetent virus-based expression vector (*e.g.*, adenovirus vector, poxvirus vector, alphavirus vector, retrovirus vector) encoding the SARS virus S protein, as a means to eliminate the need to work with infectious SARS virus.

1. Subunit SARS Vaccines Produced from SARS Virus Cultures

The SARS virus may be grown in cultured mammalian celle, such as VERO cells, then separated from the cultured cells. A SARS viral antigen, such as the S protein, can then be solubilized and separated from the SARS virus, and further isolated and purified.

In one example, the SARS virus may be produced as described in the Inactivated SARS vaccine examples, then the desired SARS antigen, such as spike protein, may be further purified from the end product using techniques known in the art.

In another example, a SARS subunit vaccine may be produced as follows. SARS virus may be produced using a desired mammalian cell line on microcarrier beads in large, controlled fermentors. For example, vaccine quality African Green Monkey kidney cells (VERO cells) at a concentration of 10^5 cells/mL are added to 60 to 75 L of CMRL 1969 media, pH 7.2, in a 150 L bioreactor containing 360 g of Cytodex-1 microcarrier beads and stirred for 2 hours. Additional CMRL 1969 is added to give a total volume of 150 L. Fetal bovine serum (FBS) is added to a final concentration of 3.5%. Glucose is added to a final concentration of 3.0 g/L and glutamine is added to a final concentration of 0.6 g/L. Dissolved oxygen, pH, agitation and temperature are controlled, and cell growth, glucose, lactate and glutamine levels are monitored. When cells are in logarithmic phases usually on days 3 to 4 reached a density of about 1.0-2.5x10^6 cells/mL, the culture medium is drained from the fermentor and 120 L of CMRL 1969, pH 7.2 (no FBS) is added and the culture stirred for 10 minutes. The draining and filling of the fermentor is usually repeated once but could be repeated up to three times. After washing the cells, the fermentor is drained and 50 L of CMRL 1969 containing 0.1% (v/v) FBS is added. The SARS virus inoculum is added at a multiplicity of infection (m.o.i.) of 0.001 to 0.01. Trypsin may be added to promote efficient infection. Additional CMRL 1969 with 0.1% FBS is added to give a final volume of 150 L. Incubation is continued at 34 C. One viral harvest is obtained from a single fermentor lot, typically at 2-7 days post-infection. Multiple harvests from a single fermentation may also be obtained.

The isolation and purification of S protein may be effected by a variety of means, as described below. For example, collecting S protein-containing flow-through from ion exchange chromatography of solubilized SARS virus envelope proteins; loading the flow through onto a hydroxyapatite matrix, and selectively eluting the S protein from the hydroxyapatite matrix. The selectively eluted S protein may be further concentrated by tangential flow ultrafiltration.

Alternatively, the isolation and purification may be effected by collecting S protein-containing flow-through from ion exchange chromatography of the solubilized SARS virus envelope proteins; loading the flow through onto a hydroxyapatite matrix and collecting an S protein-containing flow through, selectively removing detergent used in the solubilization step from the hydroxyapatite matrix flow through to provide isolated and purified S protein. The isolated and purified S protein may be subsequently concentrated by tangential flow ultrafiltration

Nucleic acid contaminants may be removed from the isolated and purified S protein by treatment with a nucleic acid degrading agent as described above in the Inactivation section. Preferably, the nucleic acid degrading agent is a nuclease, such as for example, Benzonase.

The isolated and purified S protein may be applied to a gel filtration medium and the S protein subsequently collected therefrom to separate the S protein from contaminants of other molecular weights.

Alternatively, the isolation and purification may be effected by loading S protein on a first ion-exchange medium while permitting contaminants to pass through the medium, eluting the S protein from the first ion-exchange medium, to separate the S protein from contaminants of other molecular weights. The eluted S protein is applied to a second ion-exchange medium while allowing contaminants to pass through the second ion-exchange medium. The S protein is subsequently eluted therefrom, to provide the isolated and purified S protein. The eluted S protein may be concentrated by tangential flow ultrafiltration.

Alternatively, substantially pure SARS virus S protein suitable for use as an immunogen in a subunit vaccine formulation may be prepared from infected cell lysates, such as for example using a non-denaturing detergent buffer containing 1% Triton X-100 and deoxycholate to lyse infected cells. The cell lysates are clarified by centrifugation and S protein is purified from the cell lysates by immunoaffinity purification. A monoclonal antibody against the S protein is generated and coupled to beads and a column is constructed with those beads. SARS-infected cell lysates are applied to the column, and the column is washed with PBS containing 0.1% Triton X-100. Protein bound to the column is eluted with 0.1M glycine, pH 2.5, 0.1% Triton X-100. Elution samples are buffered, such as for example, with Tris, and analyzed for the presence of protein. Fractions containing the protein are pooled and dialyzed against PBS

As discussed above, the present invention includes isolated and purified S protein of SARS virus. In one example, the virus is grown on a vaccine quality cell line, such as VERO cells, and the grown virus is harvested. The virus harvest is filtered and then concentrated typically using tangential flow ultrafiltration using a membrane of desired molecular weight cut-off and diafiltered. The virus harvest concentrate may be centrifuged and the supernatant discarded. The pellet from the centrifugation then is detergent extracted to solubilize the S protein, for example, by resuspending the pellet to the original harvest concentrate volume in an extraction buffer containing a detergent such as a non-ionic detergent including TRITON X-100.

Following centrifugation to remove non-soluble proteins, the S protein extract is purified by chromatographic procedures. The extract may first be applied to an ion exchange chromatography column such as a TMAE-fractogel or S-fractogel column equilibrated to permit the S protein to flow through while impurities are retained on the column.

Next, the flow through may be loaded onto a hydroxyapatite column, equilibrated to permit binding of the S protein to the matrix and to permit contaminants to pass from the column. The bound S protein is then eluted from the column by a suitable elutant. The resulting purified solution of S protein may be further processed to increase its purity. The eluate first may be concentrated by tangential flow ultrafiltration using a membrane of desired molecular weight cut-off. The filtrate may be contacted with a polyethylene glycol of desired molecular weight, for example, about 6000 to 8000, to precipitate the protein. Following centrifugation and discard of the supernatant, the pellet may be resuspended in PBS and dialyzed to remove the polyethylene glycol. Finally, the dialyzed solution of S protein may be sterile filtered. The sterile filtered solution may be adsorbed onto alum. The polyethylene glycol precipitation and resuspension purification step may be effected at an earlier stage of the purification operation, if desired.

Alternatively, SARS virus is recovered following growth and harvesting of the virus, and a concentrate obtained such as, for example using PEG precipitation or tangential flow filtration. The virus is contacted with detergent to solubilize the S proteins. Following centrifugation, the supernatant is recovered to further purification of the S protein and the non-soluble proteins discarded.

The supernatant is applied to an ion exchange chromatography column, such as a TMAE-fractogel or S-fractogel column, suitably equilibrated to permit retention of the S protein on the column. The S protein is eluted from the ion-exchange column under suitable conditions. The eluate then may be passed through a gel filtration column, such as a Sephacryl S-300 column, to separate the S protein from contaminants of other molecular weights. A hydroxyapatite column may be employed in place of the Sephacryl column.

The S protein may be eluted from the column to provide a purified solution of S protein. The eluate may be concentrated by tangential flow ultrafiltration using a membrane of desired molecular weight cut-off. The concentrated S protein solution then may be sterile filtered.

Alternatively, viral harvests may be concentrated by ultrafiltration and the concentrated viral harvests may be subjected to an initial purification step, for example, by gel filtration chromatography, polyethylene glycol precipitation or Cellufine sulfate chromatography. The purified virus may then be detergent extracted to solubilize the S protein. Following solubilization of the S protein, the supernatant may be loaded onto an ion-exchange column such as Cellufine sulfate chromatography column equilibrated to permit the protein to bind to the column while permitting contaminants to flow through. Similarly, a TMAE-fractogel or S-fractogel column may be used in place of the Cellufine sulfate column. The two columns also may be combined in sequential purification steps. The S protein is eluted from the columns to

provide a purified solution of the protein. This solution may be concentrated by tangential flow ultrafiltration using a membrane of desired molecular weight cut-off and diafiltered.

Specifically, in one method of S protein purification, the virus harvest concentrate is centrifuged at 28,000 x g for 30 minutes at 4 C. The supernatant is discarded and the pellet resuspended in extraction buffer consisting of 10 mM Tris-HCl, pH 7.0, 150 mM NaCl, 2% (w/v) Triton X-100 to the original harvest concentrate volume. Pefabloc is added to a final concentration of 5 mM. The suspension is stirred at room temperature for 30 minutes. The supernatant, containing the soluble S protein, is clarified by centrifugation at 28,000 x g for 30 minutes at 4 C. A TMAE--Fractogel column is equilibrated with 10 mM Tris-HCl, pH 7.0, 150 mM NaCl containing 0.02% Triton X-100. The Triton X-100 supernatant, containing the soluble S protein, is loaded directly onto the TRAE-Fractogel column. The total volume added plus 2 bed volumes of 10 mM Tris-HCl, pH 7.0, 150 mM NaCl containing 0.02% Triton X-100 are collected. The TMAE--Fractogel flow-through containing S protein is diluted 3-fold with 10 mM Tris-HCl, pH 7.0, containing 0.02% Triton X-100.

An hydroxyapatite column is equilibrated with 10 mM Tris-HCl, pH 7.0, 50 mM NaCl, 0.02% Triton X-100. After loading the TMAE flow-through, the column is washed with 2 column volumes of 10 mM Tris-HCl, pH 7.0, 50 mM NaCl , 0.02% Triton X-100 followed by 4 column volumes of 5 mM sodium phosphate, pH 7.0, 1M NaCl, 0.02% Triton X-100. The proteins are eluted with 4 column volumes of 20 mM sodium phosphate, pH 7.0, 1M NaCl , 0.02% Triton X-100. Fractions are collected based on A280 and the protein content and antigen concentrations are measured. The purified S protein is ultrafiltered by tangential flow ultrafiltration using a 300 kDa NMWL membrane.

2. Recombinant Production of Subunit SARS Vaccines

As discussed above, SARS virus proteins may be produced by recombinant expression. Host cells suitable for recombinant expression include bacterial, mammalian, insect, yeast, *etc.* Recombinant expression may be used to produce a full length SARS protein, a fragment thereof, or a fusion therewith.

Fusion peptides may be used to facilitate the expression and purification of the recombinant SARS protein. For example, recombinant production of the SARS polypeptides can be facilitated by the addition a tag protein to the SARS antigen to be expressed as a fusion protein comprising the tag protein and the SARS antigen. Such tag proteins can facilitate purification, detection and stability of the expressed protein. Tag proteins suitable for use in the invention include a polyarginine tag (Arg-tag), polyhistidine tag (His-tag), FLAG-tag, Strep-tag, c-myc-tag, S-tag, calmodulin-binding peptide, cellulose-binding domain, SBP-tag,, chitin-binding domain, glutathione S-transferase-tag (GST), maltose-binding protein, transcription termination anti-terminiantion factor (NusA), *E. coli* thioredoxin (TrxA) and protein disulfide

isomerase I (DsbA). Preferred tag proteins include His-tag and GST. A full discussion on the use of tag proteins can be found at Terpe *et al.*, "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems", Appl Microbiol Biotechnol (2003) 60:523-533.

After purification, the tag proteins may optionally be removed from the expressed fusion protein, *i.e.*, by specifically tailored enzymatic treatments known in the art. Commonly used proteases include enterokinase, tobacco etch virus (TEV), thrombin, and factor X_a.

Accordingly, the invention further includes a SARS virus subunit vaccine comprising a fusion protein. Preferably, the fusion protein comprises a first amino acid sequence encoded by a SARS virus polynucleotide sequence. SARS virus polynucleotide sequences which may encode said first amino acid sequence include one or more of the SARS virus polynucleotide sequences identified in this application and fragments thereof.

The fusion protein may comprise an amino acid sequence of a SARS virus protein or a fragment thereof. Said SARS virus protein may be selected from one or more of the group consisting of the following SARS virus proteins: P28, P65, Nsp1, Nsp2 (3CL protease), Nsp3, Nsp3, Nsp4, Nsp 5, Nsp6, Nsp 7, Nsp 8, Nsp 9 (RNA polymerase), Nsp 10 (helicase), Nsp 11, Nsp 12, Nsp 13, Spike, Orf 3, Orf 4, Envelope, Matrix, Orf 7, Orf 8, Orf 9, Orf 10, Orf 11, Nucleocapsid and Orf 13.

In one embodiment, the fusion protein comprises a first amino acid sequence comprising a SARS virus antigen or a fragment thereof. Said SARS virus amino acid sequence may comprise one or more of the T-epitope sequences identified above.

Preferably, the fusion protein comprises an amino acid sequence of a SARS virus spike protein, or a fragment thereof. Specific fragments of the spike protein which may be used in the fusion protein include the S1 domain and the S2 domain. Further fragments of the spike protein which may be used in the fusion protein include regions of each of the S1 and S2 domains, including the receptor binding region of the S1 domain, the oligomerization domain regions of the S2 domain, the leucine zipper regions of the S2 domain, the membrane anchor region of the S2 domain, the hydrophobic domain region of the S2 domain, the cystein-rich domain region of the S2 domain, and the cytoplasmic tail region of the S2 domain. (See FIGURE 19). Amino acid sequences of the Spike protein corresponding to these regions can be identified by those skilled in the art, including, for example, using the functional predictions set forth earlier in the application (predicted transmembrane helices, predicted N-terminus signaling regions, predicted coiled-coil regions, *etc.*) as well as by homology comparison to the sequences of other known Coronaviruses (See FIGURES 4F and 5).

The fusion protein may further comprise a second amino acid sequence. Said second amino acid sequence may comprise a polypeptide sequence which facilitates protein expression

or purification, preferably one of the tag sequences discussed above. Alternatively, said second amino acid sequence may comprise a second amino acid sequence from a SARS virus. Alternatively, said second amino acid sequence may comprises an amino acid sequence from another virus or bacteria, including one or more of the viruses or bacteria identified in Section I, below.

Said second amino acid sequence may comprise an amino acid sequence from another respiratory virus. Said second amino acid sequence may comprise an amino acid sequence from a virus selected from the group consisting of coronavirus, influenza virus, rhinovirus, parainfluenza virus (PIV), respiratory syncytial virus (RSV), adenovirus, and metapneumovirus.

In one embodiment, said second amino acid sequence may comprise an amino acid sequence from an adjuvant, including one or more of the adjuvants identified in section I, below.

In one embodiment, the invention includes a fusion protein comprising an amino acid sequence of a SARS virus spike protein or a fragment thereof. The fusion protein may further comprise a second amino acid sequence comprising an amino acid sequence selected from the group consisting of a second SARS virus protein, a non-SARS virus protein, a bacterial protein, and an adjuvant.

(a) Bacterial Expression of Subunit SARS Vaccines

In one embodiment, bacterial host cells are used for recombinant expression of SARS virus proteins. Bacterial host cells suitable for use in the invention include, for example, *E. coli*, *Bacillus subtilis*, and *Streptococcus spp.*

The SARS viral protein may be modified to facilitate bacterial recombinant expression. In particular, the SARS spike protein may be modified to facilitate transport of the spike protein to the surface of the bacterial host cell.

Applicants have discovered that there is strong structural homology between the SARS virus spike protein and the NadA protein of *Neisseria meningitidis*. Both proteins have an N-terminal globular "head" domain (amino acids 24-87), an intermediate alpha-helix region with high propensity to form coiled-coil structures (amino acids 88-350), and a C-terminal membrane anchor domain formed by four amphipatix transmembrane beta strands (amino acids 351-405 of NadA). In addition, a leucine zipper motive is present within the coiled-coil segment. See, FIGURE 19 depicting the SARS spike protein structure Comanducci *et al.*, "NadA, a Novel Vaccine Candidate of Neisseria meningitidis", J. Exp. Med. 195 (11): 1445-1454 (2002). In addition, a leucine zipper motif of NadA is present within the coiled-coil segment. The NadA protein also forms high molecular weight surface-exposed oligomers (corresponding to three or four monomers) anchored to meningococcal outer membrane.

When the NadA protein is expressed in *E. coli*, the full-length protein is assembled in oligomers anchored to the outer membrane of *E. coli*, similar to the way the protein is presented

in meningococcus. The NadA protein devoid of the predicted membrane anchor domain is then secreted into the culture supernatant. This secreted protein is soluble and still organized in trimers.

The invention therefore includes a fusion protein comprising an amino acid sequence of a SARS virus spike protein or a fragment thereof and a second amino acid sequence of a bacterial adhesion protein or a fragment thereof. Preferably, said adhesion protein is selected from the group consisting of NadA, YadA (of enteropathogenic *Yersinia*), and UspA2 (of *Moraxella catarrhalis*). Additional NadA-like proteins include serum resistance protein DsrA of *Haemophilus ducreyi*, the immunoglobulin binding proteins EibA, C, D, and F of *E. coli*, outer membrane protein 100 of *Actinobacillus actinomycetemcomitans*, the *saa* gene carried on the large virulence plasmid present in shiga toxigenic strains of *E. coli* (STEC), and each of the bacterial adhesion proteins described in U.K. Patent Application No. 0315022.4, filed on June 26, 2003, each of which are specifically incorporated herein by reference.

Preferably, said adhesion protein comprises NadA or a fragment thereof.

Such fusion proteins may be used to facilitate recombinant expression of immunogenic portions of SARS surface antigens, such as spike. These fusion constructs may also allow the SARS S1 and/or S2 domains to adapt to a native confirmation. These fusion proteins are also able to oligomerize and form dimers or trimers, allowing the S1 and/or S2 domains to associate and adapt conformations as in the native SARS spike protein. Further, these expression constructs facilitate surface exposure of the SARS spike protein.

The fusion proteins of the invention preferably comprise a leader peptide from a NadA like protein, preferably NadA, a polypeptide from the immunogenic "head" region of the spike protein, and a stalk region from either the NadA like protein or the Spike protein. During expression and processing of the fusion protein, one or more amino acids may be cleaved off or removed, such as, *i.e.*, the leader peptide or a membrane anchor domain.

The stalk regions facilitate oligomerization of the expression protein. Optionally, the fusion proteins of the invention further include an anchor region of a NadA like protein. This anchor region allows the expression fusion protein to anchor and assemble on the bacterial cell surface.

The fusion proteins of the invention include the following constructs:

(i) the NadA leader peptide (optionally also including the first 6 amino acids of the mature NadA protein to facilitate processing of the leader peptide and appropriate maturation of the protein) followed by the Spike S1 domain. Preferably, this construct comprises amino acids 1-29 of NadA (corresponding to the NadA leader peptide and the first 6 amino acids of the mature NadA protein, as shown in FIGURE 22 and as set forth below) followed by amino acids 14-662

of a SARS virus Spike protein (corresponding to the S1 domain, see FIGURE 19 and SEQ ID NO: 6042 and as set forth below). Specifically, construct (i) comprises SEQ ID NO: 7302.

(ii) the NadA leader peptide (optionally also including the first 6 amino acids of the mature NadA protein to facilitate processing of the leader peptide and appropriate maturation of the protein) followed by the Spike S1 domain, followed by the stalk and anchor membrane domains of NadA. Preferably, this construct comprises amino acids 1-29 of NadA (corresponding to the NadA leader peptide and the first 6 amino acids of the mature NadA protein, as shown in FIGURE 22 and as set forth below) followed by amino acids 14-662 of a SARS virus Spike protein (corresponding to the S1 domain, see FIGURE 19 and SEQ ID NO: 6042 and as set forth below) followed by amino acids 88-405 of NadA (corresponding to the stalk and the anchor membrane domains). Specifically, construct (ii) comprises SEQ ID NO: 7303.

(iii) the NadA leader peptide (optionally also including the first 6 amino acids of the mature NadA protein) followed by a SARS virus Spike S1 domain, followed by the NadA stalk domain. Preferably, this construct comprises amino acids 1-29 of NadA followed by amino acids 14-662 of a SARS virus Spike protein (corresponding to the S1 domain), followed by amino acids 88-350 of NadA (corresponding to the stalk domain). Specifically, construct (iii) comprises SEQ ID NO: 7304.

(iv) the NadA leader peptide (optionally also including the first 6 amino acids of the mature NadA protein), followed by a SARS virus Spike S1 and S2 domain (excluding the putative transmembrane region), followed by the anchor domain of NadA. Preferably, this construct comprises amino acids 1-29 of NadA, followed by amino acids 14-1195 of a SARS virus Spike protein (corresponding to S1 and S2, excluding the putative transmembrane region), followed by amino acids 351-405 of NadA (corresponding to the NadA anchor domain). Specifically, construct (iv) comprises SEQ ID NO: 7305. Alternatively, the NadA anchor domain may comprise amino acids 332 – 405 of NadA.

(v) the NadA leader peptide (optionally also including the first 6 amino acids of the mature NadA protein), followed by a SARS virus Spike S1 and S2 domain (exclusing the putative transmembrane region). Preferably, this construct comprises amino acids 1-29 of NadA, followed by amino acids 14-1195 of a SARS virus Spike protein. Specifically, construct (v) comprises SEQ ID NO: 7306.

In each of constructs (i) to (v), the first 23 amino acids are the NadA leader peptide, and the GS dipeptide at residues 679-680 arises from the insertion of a restriction enzyme site.

In constructs (i), (ii) and (iii), the NadA "head" is replaced by the Spike S1 domain, and the fusion proteins are anchored to the outer membrane of *E.coli* or secreted in the culture supernatant, respectively. In constructs (iv) and (v), the "head" and "stalk" domains of NadA are

replaced by S1 and S2 Spike domains; also in this case, the two fusion proteins are anchored to the outer membrane of *E.coli* or secreted in the culture supernatant, respectively.

Accordingly, the invention further includes a fusion protein comprising an amino acid sequence of a SARS virus spike protein or a fragment thereof and a second amino acid sequence of a bacterial adhesion protein or a fragment thereof. Preferably, amino acids corresponding to the "head" of the adhesion protein are replaced by amino acids corresponding to a SARS virus Spike S1 domain. Alternatively, the amino acids corresponding to the "head" and "stalk" domains of the bacterial adhesion protein are replaced by amino acids corresponding to the SARS virus spike protein S1 and S2 domains.

As discussed above and shown in Figure 19, the S1 domain of the Spike protein is identified as the globular receptor binding "head" region. The S1 domain of the Spike protein preferably comprises about amino acids 14-662 of SEQ ID NO: 6042. The S1 domain may comprise a shorter amino acid sequence, wherein amino acids are removed from either the N-terminal or C-terminal regions. Preferably, 3, 5, 7, 9, 13, 15, 20 or 25 amino acids are removed from either the N-terminal or C-terminal regions. The S1 domain further includes amino acid sequences having sequence identity to the S1 region of SEQ ID NO: 6042. An example of the S1 domain is SEQ ID NO: 7307:

As discussed above and shown in Figure 19, the S2 domain of the Spike protein is identified as the "stalk" region. The "stalk" region comprises oligomerization domain regions, a leucine zipper domain regions, membrane anchor regions, hydrophobic domain regions, cystein-rich domain region and a cytoplasmic tail region. The S2 domain of the Spike protein preferably excludes the transmembrane region and comprises about amino acids 663-1195 of SEQ ID NO: 6042. The S2 domain may comprise a shorter amino acid sequence, wherein amino acids are removed from either the N-terminal or C-terminal regions. Preferably, 3, 5, 7, 9, 13, 15, 20 or 25 amino acids are removed from either the N-terminal or C-terminal regions. The S2 domain further includes amino acid sequences having sequence identity to the S2 region of SEQ ID NO: 6042. An example of the S1 domain (with the transmembrane region excluded) is SEQ ID NO: 7308.

An example of the NadA protein described above is SEQ ID NO: 7309. As discussed above, the leader sequence of NadA used in the fusion protein preferably comprises about the first 29 amino acids of NadA (including a leader sequence with about 6 amino acids of the NadA head protein). Examples of such a leader sequences are set forth as SEQ ID NOS: 7310 and 7311 below. The fusion protein may use a leader sequence comprising a shorter amino acid sequence, wherein amino acids are removed from either the N-terminal or C-terminal regions. Preferably, 1, 2, 3, 4, or 5 amino acids are removed from either the N-terminal or C-terminal end of the sequence. The leader sequence used in the fusion protein may also include an amino acid

sequences having sequence identity to SEQ ID NO: 7310 or SEQ ID NO: 7311. Preferably, the leader sequence comprises SEQ ID NO: 7311.

Optionally, the fusion peptide comprises about the first 6 amino acids of the mature NadA protein to facilitate processing of the leader peptide and appropriate maturation of the protein. An examples of the first 6 amino acids of a mature NadA proteins is SEQ ID NO: 7312..

As discussed above, the stalk and anchor sequences of NadA used in the fusion protein preferably comprise about amino acids 88-405 of NadA. An example of an amino acid sequence comprising NadA stalk and anchor regions is set forth below as SEQ ID NO: 7313 below. An example of an amino acid sequence comprising a NadA stalk region (without the anchor region) is set forth as SEQ ID NO: 7314 below. An example of an amino acid sequence comprising a NadA anchor region is set forth as SEQ ID NO: 7315 below. The fusion protein may use a stalk (and/or anchor) sequence comprising a shorter amino acid sequence, wherein amino acids are removed from either the N-terminal or C-terminal regions. Preferably, 1, 2, 3, 4, 5, 6, 7, 8 or 9 amino acids are removed from either the N-terminal or C-terminal end of the sequence. The leader sequence used in the fusion protein may also include an amino acid sequences having sequence identity to the SEQ ID NO: 7313.

The fusion proteins of the invention, including those described above, may be prepared, for example, as follows. Single fragments (such as the regions described above) may be amplified by PCR using the oligonucleotide primers set forth in the Table below. ($S1_L$ refers to the Spike protein fused to the leader peptide of NadA; S2 refers to the stalk region of the Spike protein, with and without the stop codon). The oligonucleotides were designed on the basis of the DNA sequence of NadA from *N. meningitidis* B 2996 strain and of Spike from SARS virus isolate FRA1. Each oligonucleotide includes a restriction site as a tail in order to direct the cloning into the expression vector pET21b.

		SEQ ID NO:	**Restriction site**
$S1_L$	For	7316	NdeI
$S1_L$	Rev	7317	BamHI
S2	For	7318	BamHI
S2	Rev	7319	HindIII
S2-stop	Rev	7320	XhoI
$NadA_{88}$	For	7321	BamHI
$NadA_{350}$	Rev	7322	XhoI
$NadA_{332}$	For	7323	HindIII
$NadA_{405}$	Rev	7324	XhoI

The single fragments are sequentially cloned into pET21b vector, in order to express the proteins under the control of inducible T7 promoter. The S1 domain of the Spike protein fused to the leader peptide of NadA ($S1_L$) was obtained by PCR using the primers $S1_L$-For and $S1_L$-Rev. The forward oligonucleotide primer contains the NdeI restriction sequence and the

sequence coding for the leader peptide of NadA plus the first 6 aminoacids of the mature protein. The PCR fragment was cloned as a NdeI/BamHI fragment in the pET21b vector opened with the same restriction enzymes. This clone (pET-$S1_L$) was then used to sequentially clone the other different domains, as BamHI/XhoI, BamHI/HindIII or HindIII/XhoI fragments. BamHI and HindIII restriction sites introduce the aminoacids GS and KL, respectively.

The PCR amplification protocol was as follows: 200ng of genomic DNA from *Neisseria meningitidis* 2996 or 10 ng of plasmid DNA preparation (plasmid pCMVnew, containing the entire gene coding of the Spike protein), were used as template in the presence of 40μM of each oligonucletide primer, 400-800 μM dNTPs solution, 1x PCR buffer (including 1.5mM $MgCl_2$), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaQ or Invitrogen Platinum Pfx DNA polymerase).

After a preliminary 3 minute incubation of the whole mix at 95°C, each sample underwent a two-step amplification: the first 5 cycles were performed using the hybridisation temperature that excluded the restriction enzyme tail of the primer (Tm1). This was followed by 30 cycles according to the hybridisation temperature calculated for the whole length oligos (Tm2). Elongation times, performed at 68°C or 72°C, varied according to the length of the fragment to be amplified.The cycles were completed with a 10 minute extension step at 68°C or 72°C.

The amplified DNA was either loaded directly on agarose gel and the DNA fragment corresponding to the band of correct size was purified from the gel using the Qiagen™ Gel Extraction Kit, following the manufacturer's protocol.

The purified DNA corresponding to the amplified fragment and the plasmid vectors were digested with the appropriate restriction enzymes, purified using the QIAquick™ PCR purification kit (following the manufacturer's instructions) and ligation reactions were performed.

The ligation products were transformed into competent *E. coli* DH5α and screening for recombinant clones was performed by growing randomly-selected colonies and extracting the plasmid DNA using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions.

Recombinant plasmids were introduced into *E. coli* BL21(DE3) used as expression host. Single recombinant colonies were inoculated into LB + ampicillin and incubated at 37°C for 14-16 h. Bacteria were directly recovered by centrifugation (uninduced conditions) or diluted in fresh medium and grown at 37°C until OD_{600} between 0.4-0.8. Protein expression was induced by addition of 1 mM Isopropyl-1-thio-β-D-galactopyranoside (IPTG) for three hours (induced conditions).

Whole cell lysates were obtained resuspending bacteria in SDS-sample buffer 1X and boiling for 5-10 min. Equal amounts of proteins were separated using NuPAGE (Invitrogen) or

BIORAD Gel System, according to the manufacturer's instructions. Proteins were revealed by Coomassie-blue staining or transferred onto nitrocellulose membranes for western blot analysis. Western blot was performed using a rabbit polyclonal anti-serum against purified $NadA_{\Delta 351-405}$ (diluted 1:3000) and a secondary peroxidase-conjugate antibody (DAKO).

Results of the expression in *E.coli* of $S1_L$, $S1_L$-NadA and $S1_L$-$NadA_{\Delta anchor}$ are shown in FIGURES 38 and 39. Schematics of the fusion constructs are shown in FIGURE 37.

Bacterial expression of the SARS viral antigens may also be used to prepare compositions comprising outer membrane vesicles wherein said outer membrane vesicles comprise one or more SARS viral antigens.

Outer Membrane Vesicles ("OMV"), also referred to as blebs, refer to vesicles formed or derived from fragments of the outer membrane of a Gram negative bacterium. OMVs typically comprise outer membrane proteins (OMPs), lipids, phospholipids, periplasmic material and lipopolysaccharide (LPS). Gram negative bacteria often shed OMVs during virulent infections in a process known as blebbing. OMVs can also be obtained from Gram negative bacteria via a number of chemical denaturation processes, such as detergent extraction. Synthetic OMVs or liposomes, comprising a lipid bilayer and typically enclosing an aqueous core, can also be prepared with the SARS viral antigens of the invention.

The OMVs of the invention are preferably lipid vesicles comprising a lipid bilayer surrounding an aquous core. Typically the lipid vesicles are of unilamellar structure (*i.e.*, a single lipid bilayer surrounds the aquous core), although multilammelar lipid vesicles may also be used in the compositions of the invention. OMVs typically have sizes in the nanomolar to micromolar range, *e.g.*, from 1 nM to 100 μM, more typically from 10nM to 10 μM and preferably from 30 nM to 1 μM.

The OMVs of the invention are preferably prepared from gram negative bacteria. Gram negative bacteria are those bacteria that fail to resist decolorization in the commonly known Gram staining method. Gram negative bacteria are characterized by a complex multilater cell wall and often possess an outer layer polysaccharide capsule. Gram negative bacteria suitable for producing OMVs include, for example, species from *Neisseria*, *Moraxella*, *Kingella*, *Acinetobacter*, *Brucella*, *Bordetella*, *Chlamydia*, *Porphyromonas*, *Actinobacillus*, *Borelia*, *Serratia*, *Campylobacter*, *Helicobacter*, *Haemophilus*, *Escherichia*, *Legionella*, *Salmonella*, *Pseudomonas* and *Yersinia*.

The OMVs of the invention preferably comprise one or more SARS viral antigens or a fragment thereof. The SARS viral antigens may be recombinantly expressed in a Gram negative bacterial host cell and then harvested with the OMV.

Antigenic components, such as recombinantly expressed SARS viral antigens, may be located in any or all of the three main compartments of the lipid vesicles, including attached to

either the interior or exterior surface of the lipid vesicle, for example via a membrane anchor domain, or attachment to a lipid moiety; inserted into the lipid bilayer, for example where the antigenic component is itself a hydrophobic or lipid based entity; or located within the aqueous center or core of the lipid vesicle.

Synthetically prepared OMVs, or liposomes, may be used in the invention. Such liposomes may comprise a number of different lipids and fatty acids. Suitable lipids for inclusion in liposomes of the invention include but are not limited to phophatidylinositol-(4,5)-diphosphate, phosphatidylserine, phosphatidylcholine, phosphatidylethanolamine, phosphatidyglycerol, cholesterol, beta-oleolyl-gamma-palmitoyl, lipopolysaccharides and galactocerbrosides.

Suitable means for extraction of OMVs from bacterial sources include deoxycholate extraction, Tris/HCl/EDTA extraction, and lithium acetate extraction. Preferably, the extraction process comprises a physical and/or chemical means to disrupt the bacterial cell outer membrane in order to release sufficient OMVs for purification and isolation. *See, e.g.*, WO 03/051379.

The OMVs of the invention may be enriched and/or supplemented with antigenic components, such as SARS viral antigens, by methods known in the art, including, for example, direct combination *in vitro* where an energetic combination step can optionally be applied to facilitate integration of the antigenic component into a compartment of the liposome. Methods of energetic combination suitable for use in the invention include homogenization, ultrasonication, extrusion, and combinations thereof.

Preferably, the antigenic component, such as the SARS viral antigen, is recombinantly produced by the host cell from which the OMV is derived. In one embodiment, such OMVs are prepared by introducing nucleic acid sequence encoding for the SARS viral antigen into the recombinant host cell. Preferably the nucleic acid sequence encoding for the SARS viral antigen is controlled by a strong promoter sequence. Preferably, the nucleic acid sequence encoding the SARS viral antigen further comprises an outer-membrane targeting signal. For example, the nucleic acid sequence encoding the SARS viral antigen may be fused to a sequence encoding for a naturally occurring outer membrane protein of the bacterial host. Preferably, the nucleic acid sequence encoding the SARS viral antigen is fused to the signal peptide sequence of the naturally occurring outer membrane protein of the bacterial host.

Methods of preparing an optimizing OMVs for use in vaccines are disclosed in, for example Filip *et al.*, *J. Bact.* (1973) 115: 717-722; Davies *et al.*, *J. Immunol. Method* (1990) 143:215-225; and WO 01/09350.

In one embodiment, a bacterial host cell, such as *E. coli*, are transformed to express the SARS spike protein. As discussed above, the spike protein may be modified to facilitate bacterial expression and transport of the spike protein to the surface of the host cell. Each of the

Spike/NadA fusion constructs discussed above may be used in the OMV preparations of the invention. Preferably, constructs comprising the spike S1 globular head domain fused to the stalk region of NadA are used to generate OMVs. The construct may optionally include the NadA leader peptide as well as the NadA anchor peptide. Schematic diagrams of these preferred OMV constructs are depicted in FIGURE 49.

Example 6 describes one method of preparing the OMVs of the invention.

(b) Mammalian Expression of Subunit SARS Vaccine

As discussed above, mammalian host cells may be used for recombinant expression of SARS virus proteins. Mammalian host cells suitable for use in the invention include, for example, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (*e.g.*, Hep G2), Madin-Darby bovine kidney ("MDBK") cells, as well as others. Mammalian sources of cells include, but are not limited to, human or non-human primate (*e.g.*, MRC-5 (ATCC CCL-171), WI-38 (ATCC CCL-75), human embryonic kidney cells (293 cells, typically transformed by sheared adenovirus type 5 DNA), VERO cells from monkey kidneys (including, for example COS7 cells), horse, cow (*e.g.*, MDBK cells), sheep, dog (*e.g.*, MDCK cells from dog kidneys, ATCC CCL34 MDCK (NBL2) or MDCK 33016, deposit number DSM ACC 2219 as described in WO 97/37001), cat, and rodent (*e.g.*, hamster cells such as BHK21-F, HKCC cells, or Chinese hamster ovary cells (CHO cells)), and may be obtained from a wide variety of developmental stages, including for example, adult, neonatal, fetal, and embryo.

The polynucleotides encoding the SARS viral proteins may be modified to facilitate or enhance expression. For example, commercial leader sequences known in the art, such as tPA or IgK or interleukin-2, may be used in the recombinant constructs. Preferably, however, the natural SARS leader sequence is used. Use of the natural leader sequence can be used to ensure that the protein will be trafficked in human cells in the same way as during a normal viral infection, which may be advantageous *e.g.* for DNA vaccines, where antigen is expressed *in situ*.

As discussed above, tag sequences can be used in the expression constructs to facilitate purification, detection and stability of the expressed protein. Tag proteins suitable for use in the invention include a polyarginine tag (Arg-tag), polyhistidine tag (His-tag), FLAG-tag, Strep-tag, c-myc-tag, S-tag, calmodulin-binding peptide, cellulose-binding domain, SBP-tag,, chitin-binding domain, glutathione S-transferase-tag (GST), maltose-binding protein, transcription termination anti-terminiantion factor (NusA), *E. coli* thioredoxin (TrxA) and protein disulfide isomerase I (DsbA). Preferred tag proteins include His-tag and GST. A full discussion on the use of tag proteins can be found at Terpe *et al.*, "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems", *Appl Microbiol Biotechnol* (2003) 60:523-533.

After purification, the tag proteins may optionally be removed from the expressed fusion protein, *i.e.*, by specifically tailored enzymatic treatments known in the art. Commonly used proteases include enterokinase, tobacco etch virus (TEV), thrombin, and factor X_a.

One or more amino acid sequences or amino acid domains of the spike protein may be removed to facilitate mammalian recombinant expression. For instance, the entire S2 domain or the spike transmembrane region may be removed. Representative examples of some expression constructs of both full length and truncated spike glycoprotein suitable for mammalian expression are shown in FIGURE 40. Polynucleotide sequences representing each construct are shown in SEQ ID NOS 6578-6583. A description of each annotation is shown below:

Clone Name	Description	Expression Construct
nSh	natural leader sequence full length Spike histidine tag	SEQ ID NO: 6578
nS	natural leader sequence full length Spike	SEQ ID NO: 6579
nShΔTC	natural leader sequence Spike without transmembrane sequence histidine tag	SEQ ID NO: 6580
nSΔTC	natural leader sequence Spike without transmembrane sequence	SEQ ID NO: 6581
nS1h	natural leader sequence S1 domain histidine tag	SEQ ID NO: 6582
nS1	natural leader sequence S1 domain	SEQ ID NO: 6583

Cloned cDNA fragments that encompass full-length Spike coding sequences, as well as a Spike construct deleted of the transmembrane and cytoplasmic domains (TM-Cy-deleted Spike) for secretion were inserted into an expression vector pCMVIII to create nSh and nShΔTC, respectively. Both spike proteins were tagged with six histidine residues at the end of C-terminus to aid initial characterization of the expressed spike proteins. Similar sequences encoding full-length Spike or transmembrane and cytoplasmic domain deleted Spike, but without the histidine "tag" are readily substituted by one of skill in the art.

The likely locations of the expressed spike constructs was assessed by separating expressed proteins into an aqueous fraction (AF) and a detergent fraction (DF) using the procedure shown in Figure 48, with results of western blot analysis shown in Figure 43. The above described vector constructs were evaluated for expression after transfection into COS7 cells. The construct expressing the full length spike protein remained in the cell membrane while the construct expressing the truncated spike protein was located either in the cytosol (Figure 43) or secreted into the cell medium (Figure 44). As shown in Figure 43, full-length spike protein is found in DF (membrane) in an aggregated form, while the truncated protein is found in AF (cytosol) as a

monomer. As shown in Figure 44, deleted proteins (ShΔTC) are secreted, and a small fraction of full-length spike protein is detected in the medium by rabbit serum.

Recombinantly expressed spike proteins may be oligomerized. When the spike proteins are to be used in a vaccine or to generate antibodies specific to the spike protein, they are preferably oligomerized. In order to obtain oligomerized spike protein, it is preferred to maintain the transmembrane domain in the recombinant expression construct. For example, FIGURE 41 illustrates a western blot of COS7 cell lysates comparing expressed nSh and nShΔTC using both anti-his tag and rabbit anti-SARS antibodies. As shown full-length (nSh) aggregates, but the truncated (nShΔTC) spike protein does not. Antibody raised against the His-tagged protein recognizes full-length and truncated spike proteins in native and reduced forms. Rabbit antiserum recognizes spike protein only in non-reducing conditions. Spike aggregates or oligomers were present in larger amounts in the cell lysates from the expressed nSh constructs. Preferably, the oligomerized spike proteins form a homotrimer, as indicated in FIGURE 47

A further experiment, illustrated in FIGURE 42, demonstrates that the oligomerization of the expressed nSh constructs is likely due to a non-covalent linkage (and is likely not due to, for example, a disulfide bond). The oligomer dissociates into monomers at elevated temperature (80-100°C), but is stable in reducing conditions if not heated.

It is further preferred that recombinantly expressed spike proteins are glycoslyated. Tunicamycin and glycosidases were used to assess glycosylation. FIGURE 45 illustrates that glycoslation of expressed spike proteins is not affected by removal of the transmembrane domain region. Both full-length (Sh) and truncated (ShΔTC) SARS spike proteins are glycosylated.

Preferably, expression of the constructs of the invention is not toxic to the mammalian host cell. FIGURE 46 demonstrates that expression of the illustrated spike constructs is not toxic to the COS7 host cell.

Methods for transfecting, expressing, culturing, isolating and purifying recombinant proteins from mammalian cell cultures are known in the art. For example, the SARS spike constructs of the invention may be expressed in 293 cells. These cells may be cultured and transfected in static or monolayer cultures. For rapid large-scale production of SARS protein antigens in sufficient quantities for *in vitro* and *in vivo* evaluation, including immunogenicity studies, large-scale transient transfection of 293 (human embryonic kidney) cells may be used to obtain milligram quantities of the recombinant antigen(s). Alternatively, larger scale transfection of these cells may be performed with 293 cells in suspension culture. Preferably, the expressed SARS proteins are harvested from the transfected cells between 48 and 72 hours after transfection or even from 72 to 96 or more hours after transfection.

Where the host cells are transfected with truncated spike expression constructs, the expressed spike protein is secreted from the host cells and collected from the cell media. After

concentration, the spike protein may be purified from the media using, for example, GNA lectin followed by DEAE and ceramic hydroxyapatite column chromatography.

Where the host cells are transfected with full length spike expression constructs, but rather is retained within the cells, and may be purified from triton X-100 detergent extracted cells. The full-length Spike protein can then be captured on GNA lectin, followed by hydroxyapatite and SP chromatography.

Chinese Hamster Ovary (CHO) or other eukaryotic (*e.g.*, mammalian) cells that stably express the SARS viral antigens of the invention may also be derived (*e.g.* Figure 73). Preferably, the cells are CHO cells, and these constructs will comprise one or more marker or selection genes in order to select for the desired CHO cells. In one embodiment, the constructs comprise a CMV enhancer/promoter, ampicillin resistance gene, and a fused DHFR and attenuated neomycin gene for selection purposes. Stable cell lines can then be produced using the neomycin selection system in CHOK-1 cells. Selected clones can then be sequenced to verify the integrity of the insert, and transient transfections can then be performed using Trans-LT1 polyamine transfection reagent (PanVera Corp., Madison, WI) to assess the expression level and also the integrity of the expressed protein by ELISA and western blot analysis.

Methods for derivation of CHO cells stably expressing the SARS viral antigens of the invention comprise the steps of transfection and primary screening with selective medium. Optionally, these steps are followed by subcloning to assure purity of cell lines. Cell culture supernatants can be assayed using an antigen capture ELISA to quantify expression levels at all stages of selection and amplification.

For full-length Spike expression constructs, methanol fixed cells can be screened for internal expression by immunofluorescent staining using a rabbit anti-SARS antibody. Successive measurements at the T75-flask stage of expansion can be employed to assure stability of expression levels. The molecular mass and integrity of the expressed proteins can be checked by PAGE both under native and reducing and denaturing conditions, followed by immunoprobing.

In one embodiment, the pCMV3 vectors expressing SARS-CoV Spike proteins in either full-length or truncated forms is introduced into CHOK-1 cells using the Trans-LT-1 reagent. On day one, $1x10^6$ cells are plated on 100 mm dishes in non-selective F12 media + 10% Fetal Bovine Serum + 4 mM Glutamine. On day two, the cells are transfected with a DNA:LT-1 mixture and the media then replaced with complete F12 media. Twenty-four to forty-eight hours later depending on the cell density, each 100 mm dish is split to 4-6 100 mm dishes. The medium is changed to complete selective media containing Geneticin (neomycin) at 500 μg/ml. All bovine serum used in these procedures is from TSE-free sources that meet current FDA standards. Twenty-four hours later the medium is changed to complete selective medium plus

500 ug/ml neomycin. Ten to fourteen days later, individual colonies are picked and transferred to 96 well plates and cultured in complete selective medium but without G418. When approximately 80% of the wells are confluent, twenty-four hour supernatants are screened by spike capture ELISA positive clones are transferred to twenty-four well plates. For the initial expression of full length Spike protein, methanol fixed cells will be screened by immunoflourescent staining using a rabbit anti-SARS antibody. After the low expressing cell lines have been eliminated and there are less than 20-30 cell lines, capture ELISA and westerns will be used to determine the expression level after cell lysis. A portion of each cell line will be pelleted, weighed and lysed in 1% triton lysis buffer containing MOPS, NaCl and $MgCl_2$ at the same ratio of cell weight to lysis buffer. After lysis the supernatant is collected and expression level is determined. Three to four clones producing the highest levels of spike protein in correct structure and conformation will be grown in three-liter bioreactors for expansion and adaptation to low serum suspension culture conditions for scale-up.

The antigen capture ELISA assay for the SARS spike protein can be performed as described in the art. A brief description of this assay follows. 96 well flat-bottom plates (Corning, Corning, NY) are coated with 250ng per well of purified immunoglobulin obtained from rabbit sera that were immunized with inactivated SARS virus. Between steps, the plates are washed in a buffer containing 16%NaCl and 1% Triton X100. 100µL of supernatant or lysate samples (diluted in a buffer containing 100mM $NaPO_4$, 0.1% Casein, 1mM EDTA, 1% Triton X100, 0.5M NaCl and 0.01% Thiomersal, pH 7.5) are added and incubated for 2 hours at 37°C. Bound antigen is reacted against pooled SARS+ve serum or high affinity monoclonal antibody either human or mouse against SARS spike protein (1 hour incubation, 37°C) and detected using appropriate species-specific peroxidase conjugated second antibody (30 minute incubation at 37°C; TAGO, Burlingame, CA). The plates are developed for 15 minutes at room temperature using TMB substrate (Pierce, Rockford, IL) and the reaction stopped using 4N phosphoric acid. The plates are read at a wavelength of 450nm and the concentration of protein per ml sample is derived from a standard curve (OD *vs.* protein concentration) based on serial dilutions of a known concentration of recombinant spike protein.

The immunoprobing analysis can also be performed following the standard methods described elsewhere in the art. A brief description follows. 10-20 µl of the sample is analyzed on 4-20% SDS PAGE under non-reducing/ denaturing conditions with mild heating. The gels are run for 1.5-2.0 hours at 100V constant voltage. The proteins are then transferred onto nitrocellulose membranes (Millipore, Bedford, MA) for 45 min using the semidry western transfer system (BioRad, Hercules, CA) following the manufacturer's instructions. The membrane is then reacted against polyclonal anti-spike rabbit serum, followed by anti-rabbit Ig

conjugated to Alexa 688 (Molecular Probes, Oregon). The blots are scanned using an infrared imaging system (LI-Cor, Inc., Lincoln, Nebraska).

The highest expressing candidate cell lines can be screened for spike protein expression and stability in small-scale (3 liter) suspension cultures. The candidate clone can be further evaluated for level of expression as well as integrity of expressed protein after amplification, and subsequently tested for expression stability in the absence of selection. The selected clones can also be tested for maintenance of the DNA sequence integrity of the integrated SARS spike protein gene. To quickly monitor the expression levels in small flask (T25 or T75) and in the three liter evaluation cultures, a lectin-based process (Gluvanthus Nivalis lectin) may be used to isolate SARS spike protein to a degree of purity that allows semi-quantitation and characterization of the protein in CHO supernatant. For full-length spike protein, it will be obtained from triton X-100 detergent extracted cells. Full-length Spike protein will be then captured on GNA lectin, followed by hydroxyapatite and SP chromatograph. Eluted protein is then characterized by: 1) polyacrylamide gel electrophoresis (PAGE) and Coomassie staining, 2) Immunoprobing with anti-SARS rabbit sera, 3) structural characterization using size exclusion chromatography (SEC), as well as mass spec analysis using MALDI-TOF.

Routes and methods of immunization of the vaccines of the invention are discussed in more detail in a section below. Examples 7 to 9 illustrate sample immunization protocols for the recombinant spike proteins.

Vaccine testing

Prior to human administration, it is normal to test vaccines in animal models. A mouse model of SARS coronavirus infection is known (Subbarao *et al.* (2004) *J Virol* 78:3572-77), and other animals that may be used as models of infection and/or disease include ferrets and monkeys. Thus the invention provides a non-human animal that is infected by the SARS coronavirus, wherein the animal is preferably a ferret or a primate (*e.g.* a monkey or a macaque). The animal may be gnotobiotic. The animal is preferably not a cat (*Felis domesticus*). The animal may or may not display SARS disease symptoms *e.g.* ferrets (*Mustela furo*) show prominent pulmonary pathology after infection. See: Martina *et al.* (2003) *Nature* 425:915.

E. Polynucleotides encoding the SARS Antigens of the Invention

The invention includes polynucleotides encoding for the SARS antigens of the invention. In addition, the invention includes polynucleotides which have been optimized for recombinant production (*e.g.* codon optimization) of the SARS antigens of the invention, including polynucleotides encoding for each of the SARS fusion constructs discussed above.

F. Viral vector or Viral Particle delivery of the SARS Antigens of the Invention

The antigens of the invention may be expressed *in vivo* or *in vitro* by polynucleotides encoding the antigens. Expression and delivery of the polynucleotides of the invention may be facilitated via viral vectors and/or viral particles.

Gene-based delivery systems derived from viruses, such as alphaviruses, are useful for the *ex vivo* and *in vivo* administration of heterologous genes, including one or more SARS genes, having therapeutic or prophylactic applications. These systems can also be used for the production of recombinant proteins derived from the SARS virus in cultured cells. Gene-based delivery systems of the invention include viral vectors (*e.g.*, adenovirus vector, poxvirus vector, alphavirus vector) and non-viral nucleic acid vectors (*e.g.*, DNA, RNA) encoding one or more SARS virus antigens. Polynucleotides encoding SARS virus antigen(s) are incorporated into the gene-based vaccines individually or in combination (*e.g.*, as bicistronic constructs).

1. Alphavirus

Alphaviruses are members of *Togaviridae* family and share common structural and replicative properties. Sindbis virus (SIN) is the prototype virus for the molecular study of other alphaviruses, and together with Venezuelan equine encephalitis virus (VEE) and Semliki Forest virus (SFV), are the most widely utilized alphaviruses being developed into expression vectors for heterologous genes (Schlesinger and Dubensky (1999) *Curr Opin. Biotechnol. 10*:434-439; Schlesinger (2001) Expert Opin. Biol. Ther. 1:177-91).

Alphaviruses possess a relatively small single-stranded RNA genome of positive polarity, which is approximately 12 kb in length, capped and polyadenylated. The RNA interacts with viral capsid protein monomers to form nucleocapsids, which in turn, are surrounded by a host cell-derived lipid envelope from which two viral glycoproteins, E1 and E2, protrude forming "spike" trimers of heterodimeric subunits. Two open reading frames (ORFs) encode as polyproteins the enzymatic nonstructural replicase proteins (5' ORF) and the virion structural proteins (3' ORF). The structural polyprotein is translated from a highly abundant subgenomic mRNA, which is transcribed from a strong internal alphavirus promoter (Strauss and Strauss (1994) *Microbiol. Rev.* 58:491-562). Replication of the genome occurs exclusively within the host cell cytoplasm as RNA.

The most common alphavirus expression vectors have exploited both the positive-stranded nature and modular organization of the RNA genome. These vectors, termed "replicons" due to their property of self-amplification, permit insertion of heterologous sequences in place of the structural polyprotein genes, while maintaining the 5'- and 3'-end *cis* replication signals, the nonstructural replicase genes, and the subgenomic junction region promoter (Xiong *et al.* (1989) *Science 243*:1188-1191; Liljestrom (1991) *Bio/Technology 9*:1356-1361). Chimeric alphavirus vectors (and particles) from sequences derived from divergent virus families have also been

described. (see, for example United States patent application serial number 09/236,140; see also, US Patents 5,789,245, 5,842,723, 5,789,245, 5,842,723, and 6,015,694; as well as WO 95/07994, WO 97/38087 and WO 99/18226). Co-owned International Publication WO 02/099035, published December 12, 2002 and incorporated by reference in its entirety herein, describes chimeric alphavirus molecules and modified alphavirus molecules having modified Biosafety Levels.

The absence of structural protein genes renders alphavirus replicon vectors defective, in that RNA amplification and high-level heterologous gene expression occurs within the target cell, but cell-to-cell spread of vector is not possible due to the inability to form progeny virions. Through the years, several synonymous terms have emerged that are used to describe alphavirus replicon particles. These terms include recombinant viral particle, recombinant alphavirus particle, alphavirus replicon particle and replicon particle. However, as used herein, these terms all refer to a virion-like unit containing an alphavirus-derived RNA vector replicon. Moreover, these terms may be referred to collectively as vectors, vector constructs or gene delivery vectors.

Packaging of replicon RNA into particles can be accomplished by introducing the replicon RNA into permissive cells (*e.g.*, RNA or DNA transfection, or particle infection) that also contain one or more structural protein expression cassettes or "defective helper" constructs encoding the alphavirus structural proteins. These structural protein encoding constructs may themselves be introduced into the cells by transfection of either RNA or DNA, and most commonly retain the native alphavirus subgenomic promoter, as well as 5'- and 3'-end *cis* signals for co-amplification with the replicon, but are devoid of any replicase genes and the RNA packaging signal (Liljestrom (1991) *Bio/Technology 9*:1356-1361; Pushko *et al.* (1997) *Virology 239*:389-401; Polo *et al.* (1999) *PNAS 96*:4598-4603). Permanent cell lines that are stable transformed with constructs expressing the alphavirus structural proteins (*e.g.*, packaging cell lines) offer a means to avoid transient transfection production methods (Polo *et al.* (1999) *PNAS 96*:4598-4603).

The present invention includes compositions and methods for the production of replication defective viral vector particles (*e.g.*, alphavirus replicon particles) for use in the *ex vivo* and *in vivo* administration of heterologous genes encoding proteins having therapeutic or prophylactic application, including genes encoding for one or more SARS viral antigens.

In one aspect, the invention includes a method of producing replication defective viral vector particles (*e.g.*, alphavirus replicon particles) comprising the steps of introducing at least one nucleic acid molecule comprising a viral vector (*e.g.*, alphavirus replicon RNA) into immortalized cells of the present invention, under conditions that allow for complementation of the viral vector (*e.g.*, alphavirus replicon RNA) and production of viral vector particles (*e.g.*, alphavirus replicon particles), and isolating the viral vector particles from the cells or cell culture

supernatants. In certain embodiments, the immortalized cells are grown in suspension, for example PERC.6 cells. In other embodiments, the methods are performed in large-scale volumes, for example, liter volumes or greater, such as for example in roller bottles, large flasks, Nunc Cell Factories, Corning Cell Cubes, fermentation vessels, etc).

In certain embodiments, the viral vector is an alphavirus replicon RNA that requires complementation by providing one or more alphavirus structural proteins in trans, within the immortalized cell. In such instances, the methods of complementation to produce alphavirus replicon particles may involve the introduction of one or more nucleic acids (*e.g.*, RNA, DNA) encoding said alphavirus structural protein(s) (*e.g.*, capsid and/or envelope glycoproteins) into the immortalized cells, either transiently or stably, and either concurrent with or prior to the introduction of the alphavirus replicon RNA. In certain embodiments, the alphavirus replicon RNA is introduced into the cell by transfection an in vitro transcribed RNA. In other embodiments, the alphavirus replicon RNA is introduced into the cell by transfection of a DNA (*e.g.*, ELVIS), which is capable of transcribing within the cell, the replicon RNA. In yet other embodiments, the alphavirus replicon RNA is introduced into the cell by infection with a seed stock of alphavirus replicon particles. In certain embodiments, the nucleic acids encoding said alphavirus structural protein(s) are defective helper RNA or are DNA that can transcribe within the cell defective helper RNAs.

As discussed herein, "alphavirus RNA replicon vector", "RNA replicon vector", "replicon vector" or "replicon" refers to an RNA molecule that is capable of directing its own amplification or self-replication *in vivo*, within a target cell. To direct its own amplification, the RNA molecule should encode the polymerase(s) necessary to catalyze RNA amplification (*e.g.*, alphavirus nonstructural proteins nsP1, nsP2, nsP3, nsP4) and also contain *cis* RNA sequences required for replication which are recognized and utilized by the encoded polymerase(s). An alphavirus RNA vector replicon should contain the following ordered elements: 5' viral or cellular sequences required for nonstructural protein-mediated amplification (may also be referred to as 5' CSE, or 5' *cis* replication sequence, or 5' viral sequences required in *cis* for replication, or 5' sequence which is capable of initiating transcription of an alphavirus), sequences which, when expressed, code for biologically active alphavirus nonstructural proteins (*e.g.*, nsP1, nsP2, nsP3, nsP4), and 3' viral or cellular sequences required for nonstructural protein-mediated amplification (may also be referred as 3' CSE, or 3' viral sequences required in *cis* for replication, or an alphavirus RNA polymerase recognition sequence). The alphavirus RNA vector replicon also should contain a means to express one or more heterologous sequence(s), such as for example, an IRES or a viral (*e.g.*, alphaviral) subgenomic promoter (*e.g.*, junction region promoter) which may, in certain embodiments, be modified in order to increase or reduce viral transcription of the subgenomic fragment, or to decrease homology with

defective helper or structural protein expression cassettes, and one or more heterologous sequence(s) to be expressed. Preferably the heterologous sequence(s) comprises a protein-encoding gene, which is the 3' proximal gene within the vector replicon. And preferably the replicon further comprises a polyadenylate tract.

As discussed herein, "recombinant Alphavirus Particle", "alphavirus replicon particle" and "replicon particle" refers to a virion-like unit containing an alphavirus RNA vector replicon. Generally, the recombinant alphavirus particle comprises one or more alphavirus structural proteins, a lipid envelope and an RNA vector replicon. Preferably, the recombinant alphavirus particle contains a nucleocapsid structure that is contained within a host cell-derived lipid bilayer, such as a plasma membrane, in which one or more alphaviral envelope glycoproteins (*e.g.*, E2, E1) are embedded. The particle may also contain other components (*e.g.*, targeting elements such as biotin, other viral structural proteins or portions thereof, hybrid envelopes, or other receptor binding ligands), which direct the tropism of the particle from which the alphavirus was derived. Generally the interaction between alphavirus RNA and structural protein(s) necessary to efficiently form a replicon particle or nucleocapsid may be an RNA-protein interaction between a capsid protein and a packaging signal or packaging sequence contained within the RNA.

"Alphavirus packaging cell line" refers to a cell which contains one or more alphavirus structural protein expression cassettes and which produces recombinant alphavirus particles (replicon particles) after introduction of an alphavirus RNA vector replicon, eukaryotic layered vector initiation system, or recombinant alphavirus particle. The parental cell may be of mammalian or non-mammalian origin. Within preferred embodiments, the packaging cell line is stably transformed with the structural protein expression cassette(s).

"Defective helper RNA" refers to an RNA molecule that is capable of being amplified and expressing one or more alphavirus structural proteins within a eukaryotic cell, when that cell also contains functional alphavirus nonstructural "replicase" proteins. The alphavirus nonstructural proteins may be expressed within the cell by an alphavirus RNA replicon vector or other means. To permit amplification and structural protein expression, mediated by alphavirus nonstructural proteins, the defective helper RNA molecule should contain 5'-end and 3'-end RNA sequences required for amplification, which are recognized and utilized by the nonstructural proteins, as well as a means to express one or more alphavirus structural proteins. Thus, an alphavirus defective helper RNA should contain the following ordered elements: 5' viral or cellular sequences required for RNA amplification by alphavirus nonstructural proteins (also referred to elsewhere as 5' CSE, or 5' *cis* replication sequence, or 5' viral sequences required in *cis* for replication, or 5' sequence which is capable of initiating transcription of an alphavirus), a means to express one or more alphavirus structural proteins, gene sequence(s) which, when expressed,

codes for one or more alphavirus structural proteins (*e.g.*, C, E2, E1), 3' viral or cellular sequences required for amplification by alphavirus nonstructural proteins (also referred to as 3' CSE, or 3' viral sequences required in *cis* for replication, or an alphavirus RNA polymerase recognition sequence), and a preferably a polyadenylate tract. Generally, the defective helper RNA should not itself encode or express in their entirety all four alphavirus nonstructural proteins (nsP1, nsP2, nsP3, nsP4), but may encode or express a subset of these proteins or portions thereof, or contain sequence(s) derived from one or more nonstructural protein genes, but which by the nature of their inclusion in the defective helper do not express nonstructural protein(s) or portions thereof. As a means to express alphavirus structural protein(s), the defective helper RNA may contain a viral (*e.g.*, alphaviral) subgenomic promoter which may, in certain embodiments, be modified to modulate transcription of the subgenomic fragment, or to decrease homology with replicon RNA, or alternatively some other means to effect expression of the alphavirus structural protein (*e.g.*, internal ribosome entry site, ribosomal readthrough element). Preferably an alphavirus structural protein gene is the 3' proximal gene within the defective helper. In addition, it is also preferable that the defective helper RNA does not contain sequences that facilitate RNA-protein interactions with alphavirus structural protein(s) and packaging into nucleocapsids, virion-like particles or alphavirus replicon particles. A defective helper RNA is one specific embodiment of an alphavirus structural protein expression cassette.

Alphavirus for use in the invention may be grown in any one of the cell lines discussed above as suitable for the SARS virus.

Alphavirus replicon particles may be produced according to the present invention by using the above cell lines (*e.g.*, immortalized cell lines) and a variety of published and accepted alphavirus vector methodologies. Such methodologies include, for example, transient packaging approaches, such as the co-transfection of *in vitro* transcribed replicon and defective helper RNA(s) (Liljestrom, *Bio/Technology 9*:1356-1361, 1991; Bredenbeek *et al.*, *J. Virol. 67*:6439-6446, 1993; Frolov *et al.*, *J. Virol. 71*:2819-2829, 1997; Pushko *et al.*, *Virology 239*:389-401, 1997; US Patents 5,789,245 and 5,842,723) or co-transfection of plasmid DNA-based replicon and defective helper construct(s) (Dubensky *et al.*, *J. Virol. 70*:508-519, 1996), as well as introduction of alphavirus structural protein expression cassettes (*e.g.*, DNA-based defective helper) into immortalized cell lines of the present invention to create stable packaging cell lines (PCL) (Polo *et al.*, *PNAS 96*:4598-4603, 1999; US Patents 5,789,245, 5,842,723, 6,015,694; WO 97/38087, WO 99/18226, WO 00/61772, and WO 00/39318). Stable packaging cell lines may then be utilized for alphavirus replicon particle production. The PCL may be transfected with *in vitro* transcribed alphavirus replicon RNA, transfected with a plasmid DNA-based replicon (*e.g.*, ELVIS vector), or infected with a seed stock of alphavirus replicon particles, and then incubated under conditions and for a time sufficient to produce progeny alphavirus replicon particles in the

culture supernatant. In addition, progeny replicon particles can subsequently be passaged in additional cultures of naïve PCL by infection, resulting in further expansion and commercial scale preparations. Importantly, by using defective helper RNA or stable PCL based on the "split" structural gene configuration, these replicon particle stocks may be produced free from detectable contaminating RCV.

Following harvest, crude culture supernatants containing the chimeric alphavirus replicon particles may be clarified by passing the harvest through a filter (*e.g.*, 0.2 uM, 0.45 uM, 0.65 uM, 0.8 uM pore size). Optionally, the crude supernatants may be subjected to low speed centrifugation prior to filtration to remove large cell debris. Within one embodiment, an endonuclease (*e.g.*, Benzonase, Sigma #E8263) is added to the preparation of alphavirus replicon particles before or after a chromatographic purification step to digest exogenous nucleic acid. Further, the preparation may be concentrated prior to purification using one of any widely known methods (*e.g.*, tangential flow filtration). Crude or clarified alphavirus replicon particles may be concentrated and purified by chromatographic techniques (*e.g.*, ion exchange chromatography, size exclusion chromatography, hydrophobic interaction chromatography, affinity chromatography), such as those described in WO01/92552, incorporated by reference in its entirety herein. Two or more such purification methods may be performed sequentially.

EXAMPLE OF ALPHAVIRUS REPLICON PARTICLES ENCODING SARS VIRUS SPIKE (S) ANTIGEN

The invention includes compositions and methods for the production of replication defective viral vector particles (*e.g.*, alphavirus replicon particles) for use in the ex vivo and in vivo administration of heterologous genes encoding proteins having therapeutic or prophylactic application, including genes encoding for one or more SARS viral antigens.

The following example illustrates a method of preparing alphavirus replicon particles encoding SARS virus spike (s) antigen.

The SARS virus spike gene can be incorporated into alphavirus replicon particles derived from a variety of alphavirus, such as Sindbis virus, Semliki Forest virus (US 5739026), Venezuelan equine encephalitis virus (US 6531135), and replicon particle chimeras derived from more than one alphavirus (US 6376236, WO 02/99035). In addition, the SARS virus spike gene can be incorporated in its entirety (encoding full-length spike protein) or in a modified form that includes, for example, sequence deletions or truncations, such that the encoded a spike protein is of less than full-length (*e.g.*, C-terminal truncation of one or more (*e.g.* at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30 *etc.*) amino acids, deleted of transmembrane region and cytoplasmic tail).

For example, the spike gene may be cloned as a full-length gene into the VCR-chim2.1 vector (WO 02/99035) by standard RT-PCR conditions or by standard subcloning from one of the other plasmids described herein, using commercially available restriction endonucleases. For

the reverse transcription step in standard RT-PCR, the Superscript pre-amplification kit (Invitrogen™) and the primer SEQ ID NO: 7325 (sp-RT-R) are used:

For the amplification step, the cDNA polymerase advantage kit (Clonetech) and two primers Sp-F-BbvCI (SEQ ID NO: 7326) and Sp-R-NotI (SEQ ID NO: 7327) are used:

The forward primer is designed to contain the ccacc sequence (Kozak, 1991 *JBC* 19867-70) in front of the ATG codon to optimize translation efficiency of the spike gene. Also, the forward primer contains the BbvCI restriction site and the reverse primer contains the NotI restriction site for subsequent cloning of the PCR amplified gene.

The PCR product is purified using the QIAquick Nucleotide Removal kit (QIAgen), digested with BbvCI and NotI, gel purified with QIAquick Gel Extraction kit (QIAgen), and ligated to plasmid VCR-Chim2.1 pre-digested with the same enzymes. Clones containing the SARS spike sequence are verified by sequencing and the new construct is called VCR-Chim2.1-SARSspike.

To generate VEErep/SINenv-SARSspike replicon particles the plasmids VCR-Chim2.1-SARSspike, VCR-DH-Scap (WO 02/99035), and VCR-DH-Sglydl160 (WO 02/99035) are linearized with the restriction enzyme PmeI and used for *in vitro* transcription as described previously (Polo *et al.* 1999, PNAS 96: 4598-603; WO02/99035). The transcripts are co-transfected into BHK cells as previously described (Polo *et al.*, 1999, *ibid.*; WO02/99035). The transfected cells are incubated at 34 ^{0}C, the supernatants collected at 20 and 30 hrs post-electroporation, clarified by centrifugation, and purified by chromatography as previously described (WO 01/92552).

Expression of the SARS spike protein from the replicon particle vector is verified by infecting BHK cells overnight with purified VEErep/SINenv-SARSspike or VEErep/SINenv-GFP (WO 02/99035) replicon particles. In addition, BHK cells also were transfected in parallel with in vitro transcribed VCR-Chim2.1-SARSspike replicon RNA. At 16 hrs post-infection and transfection cells are lysed and a sample of the lysate analyzed by western blot using an antibody that recognizes SARS virus spike protein. The proteins on the gel are stained or transferred to a membrane for Western blot analysis with sera from convalescent patients or alternatively murine or rabbit antisera generated against SARS virus. VEErep/SINenv-SARSspike replicon particles are administered to the vaccine recipient (*e.g.*, rodent, non-human primate, human) as described elsewhere in the present invention.

Figure 67 shows data from western blot analysis performed under non-reducing conditions, using a SARS virus specific rabbit polyclonal antisera. The western data demonstrate that not only is SARS spike protein expressed in cells infected with alphavirus replicon particles or transfected with replicon RNA, but the predominant form of spike is that of a homotrimer (Fig.67A). Similar homotrimeric association of the spike protein was observed in western blots

of SARS virions purified from SARS virus infected VERO cell supernatants, and this homotrimer is heat labile, as indicated by the dissociation into monomeric forms at 80°C and 100°C (Fig.67B).

To further characterize SARS Spike protein expression and processing following expression from alphavirus replicon vectors, BHK-21 cells were infected with alphavirus replicon particles expressing the full-length Spike. At 6 hr post-infection with an MOI of 5, infected cells were labeled for 1 hr with L-[^{35}S]methionine/cysteine and chased for the indicated time. The [^{35}S]-labeled spike protein was immunoprecipitated by anti-SARS rabbit serum and digested with Endo-H. Both digested and undigested proteins were analysed by 4% polyacrylamide-SDS PAGE under reducing conditions. As shown in Figure 55, the full-length spike protein is synthesized as an Endo-H sensitive high mannose glycoprotein (gp170, an ER form) that undergoes modification to an Endo-H resistant glycoprotein with complex oligosaccharides (gp180, a Golgi form). The conversion of gp170 into the gp180 form takes place within 2 hr.

Alphavirus replicon particles expressing one or more SARS proteins (e.g., VEErep/SINenv-SARSspike replicon particles) are administered to the vaccine recipient in order to induce a SARS specific immune response (*e.g.*, rodent, ferret, non-human primate, human) as described elsewhere in the present invention. Immunization may be performed through a variety of routes, including for example, intramuscular, subcutaneous, intradermal, and intranasal. In addition, the alphavirus replicon particles may be used alone or in combination (*e.g.*, "prime-boost") with other vaccine approaches of the present invention, or alternatively the alphavirus replicon particles may co-express antigen from other respiratory pathogens or be co-administered in combination with alphavirus replicon particles expressing antigens from other respiratory pathogens (*e.g.*, influenza virus, parainfluenza virus, respiratory syncytial virus, human metapneumovirus). For example, the induction of anti-spike protein antibodies in animals immunized IM with VEErep/SINenv-SARSspike replicon particles was demonstrated in mice (Figure 68). These mouse studies also included addtional vaccine groups for comparison, including the inactivated SARS virus and recombinant truncated spike protein vaccines describe elsewhere herein, as well as plasmid DNA used as a prime, followed by alphavirus replicon particles as a boost. The data clearly show very potent immune responses for all vaccine groups, including the alphavirus replicon particle group. It should be noted that the level of antibody induced by the inactivated SARS virus vaccine used in these experiments has been shown to be protective in a SARS virus animal challenge model.

Similarly, genes encoding other SARS virus antigens (*e.g.*, nucleocapsid protein, membrane glycoprotein) are cloned into alphavirus replicon vectors, either individually or in

combination, to generate alphavirus replicon particles according to the teachings of the present invention and using standard molecular biology techniques..

EXAMPLE OF ALPHAVIRUS-BASED PLASMID DNA EXPRESSING SARS VIRUS SPIKE (S)

The invention includes preparation of plasmid DNA expressing a SARS virus antigen for prophylactic or therapeutic immunization against SARS virus infection. In one embodiment, the SARS viral antigen is a spike (S) protein. In one embodiment, the plasmid DNA is alphavirus-based.

The following example illustrates one method for preparing an alphavirus-based plasmid DNA expressing SARS virus spike (S).

SARS spike gene can be delivered using any of the alphavirus-based plasmid DNA replicons such as ELVS (Dubensky et al, 1996 J Virol. 70: 508-19), SINCP (WO 01/81609), or VCP (PCT WO 02/99035).

For example, the SARS spike gene is cloned into SINCP using the standard RT-PCR techniques. The oligo Sp-RT-R is used for the reverse transcription step with the Superscript pre-amplification kit (Invitrogen). For the amplification step, the cDNA polymerase advantage kit (Clonetech) with the Sp-R-NotI and Sp-F-XhoI (SEQ ID NO: 7328) primers is used.

The Sp-F-XhoI primer was designed to contain the ccacc sequence in front of the ATG codon to optimize translation efficiency (Kozak 1991, ibid) of the spike gene. Also, the primer contains the XhoI restriction site for the subsequent cloning of the PCR amplified gene.

The PCR product is purified using the QIAquick Nucleotide removal kit, digested with XhoI and NotI, gel purified with QIAquick Gel Extraction kit, and ligated to plasmid SINCP pre-digested with the same enzymes. Clones containing the SARS spike sequence are verified by sequencing and the new construct is called SINCP-SARSspike.

Expression of the SARS spike gene is verified by transient transfection of BHK cells with 2μg of either plasmid DNA SINCP-SARSspike or SINCP pre-incubated for 5 minutes with 5 μl of TransIT Polyamine reagent (Mirrus) in low serum medium Optimem (Life Technologies). At 48 hrs pos-transfection cells are lysed and a sample of the lysate is run on 8% SDS-PAGE. The proteins on the gel are either stained or transferred to a membrane for Western blot analysis with sera from convalescent patients, or alternatively with sera from mouse or rabbits.

SINCP-SARSspike plasmid replicon is administered to the vaccine recipient (*e.g.*, rodent, non-human primate, human) as a formulated or unformulated plasmid vaccine, alone or in combination (*e.g.*, "prime-boost") with other vaccines of the present invention, as described elsewhere herein.

Similarly, genes encoding other SARS virus antigens (*e.g.*, nucleocapsid protein, membrane glycoprotein) are cloned into alphavirus plasmid replicon vectors.

2. Plasmid Expression Vectors

EXAMPLE OF PLASMID DNA EXPRESSING SARS VIRUS SPIKE (S)

The following example illustrates a method for preparing plasmid DNA expressing SARS virus spike (s).

The SARS virus spike antigen also may be delivered using other plasmid DNA expression vectors (sometimes referred to as "conventional" DNA vaccines), based on a polymerase II promoter, such as, for example, a CMV promoter. A DNA vaccine of the spike antigen gene induces an antibody response in mice (Zhao *et al.* (2004) *Acta Biochim et Biophysica Sinica* 36:37-41), and has been found to induce viral neutralization and protective immunity in mice (Yang *et al.* (2004) *Nature* 428:561-564), particularly when truncated at the C-terminus.

For example, the SARS spike gene is cloned into pCMVKm2 (Zur Megede *et al.*, J.Virol., 74:2628-2635, 2000; SEQ ID NO: 9923) using standard RT-PCR techniques. The oligo Sp-RT-R is used for the reverse transcription step with the Superscript pre-amplification kit (Invitrogen). For the amplification step, the cDNA polymerase advantage kit (Clonetech) is used with primers Sp-F-EcoRI (SEQ ID NO: 7329) and Sp-R-XbaI (SEQ ID NO: 7330).

The forward primer was designed to contain the CCACC (SEQ ID NO: 7331) sequence in front of the ATG codon to optimize translation efficiency (Kozak 1991, ibid.) of the spike gene. Also, the forward primer contains the EcoRI restriction site and the reverse primer contains the XbaI restriction site for the subsequent cloning of the PCR amplified gene.

The PCR product is purified using the QIAquick Nucleotide Removal kit, digested with XhoI and NotI, gel purified with QIAquick Gel Extraction kit, and ligated to plasmid pCMVKm2 pre-digested with the same enzymes. Clones containing the SARS spike sequence are verified by sequencing and the new construct is called pCMVKm2-SARSspike.

Expression of the SARS spike gene is verified by transient transfection of BHK or 293 cells with 2μg of either plasmid DNA pCMVKm2-SARSspike or pCMVKm2 pre-incubated for 5 minutes with 5 μl of TransIT Polyamine reagent (Mirrus) in low serum medium Optimem (Life Technologies). At 48 hrs pos-transfection cells are lysed and a sample of the lysate is run on 8 % SDS-PAGE. The proteins on the gel are either stained or transferred to a membrane for Western blot analysis with sera from convalescent patients, or alternatively using mouse or rabbit antisera.

Plasmid pCMVKm2-SARSspike is administered to the vaccine recipient (*e.g.*, rodent, non-human primate, human) as a formulated or unformulated plasmid vaccine, as described elsewhere in the present invention.

Similarly, genes encoding other SARS virus antigens (*e.g.*, nucleocapsid protein, membrane glycoprotein) are cloned into plasmid expression vectors

3. Virus-Like Particles comprising SARS antigens

The SARS viral antigens of the invention may be formulated into Virus Like Particles ("VLPs"). The invention thus includes virus-like particles (or VLPs) comprising one or more SARS viral antigens. Preferably, the VLPs comprise one or more SARS viral antigens selected from the group consisting of Spike (S), nucleocapsid (N), membrane (M) and envelope (E). Preferably, the VLPs comprise at least M and E.

The VLPs of the invention comprise at least one particle-forming polypeptide. Said particle-forming polypeptide is preferably selected from a Coronavirus structural protein. In one embodiment, the particle-forming polypeptide is selected from one or more SARS viral antigens. In another embodiment, the particle-forming polypeptide is selected from the structural protein of a non-SARS Coronavirus, such as, for example, Mouse Hepatitis Virus.

VLPs can be formed when viral structural proteins are expressed in eukaryotic or prokaryotic expression systems. Upon expression, the structural proteins self-assemble to form particles. Alternatively, viral structural proteins may be isolated from whole virus and formulated with phospholipids. Such viral structural proteins are referred to herein as "particle-forming polypeptides". VLPs are not infectious because no viral genome is present, however, these non-replicating, virus capsids mimic the structure of native virions.

Due to their structure, VLPs can display a large number of antigenic sites on their surface (similar to a native virus). VLPs offer an advantage to live or attenuated vaccines in that they are much safer to both produce and administer, since they are not infectious. VLPs have been shown to induce both neutralizing antibodies as well as T-cell responses and can be presented by both class I and II MHC pathways.

Previous work creating VLPs from coronavirus indicates that E and M proteins along may be sufficient for coronavirus VLP formation. *See* Fischer *et al., J. Virol.* (1998) 72:7885-7894 and Vennema *et al. EMBO J.* (1996) 15:2020-2028.

Chimeric VLPs comprising particle-forming polypeptides or portions thereof from non-SARS Coronaviruses are also included in the invention. Such particle-forming polypeptides may comprise a full length polypeptide from a non-SARS Coronavirus. Alternatively, a particle-forming fragment may be used.

In one embodiment, a fragment of a non-SARS particle-forming polypeptide and a fragment of a SARS viral antigen are fused together. For instance, such chimeric polypeptides may comprise the the endodomain and transmembrane domain of a non-SARS particle-forming polypeptide and the ectodomain of a SARS viral antigen. In one example, the VLPs of the invention comprise a chimeric spike protein comprising an endodomain and transmembrane domain of the spike protein of Mouse Hepatitis Virus (MHV) and the chimeric spike protein further comprises the ectodomain of the SARS spike protein. Such VLPs may further comprise

Coronavirus M and E proteins. Said M and E proteins may be selected from any coronavirus, including Mouse Hepatitis Virus (MHV) or SARS. Sample sequences of S, M and E proteins of MHV are included in the figures, *supra*.

Chimeric spike proteins derived from the ectodomain of feline infectious peritonitis virus (FIPV) spike protein fused to the endo and transmembrane domains of MHV spike protein have been previously disclosed. See WO 98/49195 and WO 02/092827. In these chimeric VLP structures, the capsid structure of the VLPs is formed by the M and E protein of MHV. The chimeric spike protein provides for the surface exposure of the ectodomain of the FIPV spike protein.

As used herein, the term "virus-like particle" or "VLP" refers to a non-replicating, empty virus shell. VLPs are generally composed of one or more viral proteins, such as, but not limited to those proteins referred to as capsid, coat, shell, surface and/or envelope proteins, or particle-forming polypeptides derived from these proteins. VLPs can form spontaneously upon recombinant expression of the protein in an approrpirate expression system. Alternatively, viral structural proteins may be isolated from whole virus and formulated with phospholipids. Methods for producing particular VLPs are known in the art and discussed more fully below. The presence of VLPs in a composition can be detected using conventional techniques known in the art, such as by electron microscopy, x-ray crystallography, and the like. See, *e.g.*, Baker *et al.*, *Biophys. J.* (1991) 60:1445-1456; Hagensee *et al.*, *J. Virol.* (1994) 68:4503-4505. For example, cryoelectron microscopy can be performed on vitrified aqueous samples of the VLP preparation in question, and images recorded under appropriate exposure conditions.

The phrase "particle-forming polypeptide" includes a full-length or near full-length viral protein, as well as a fragment thereof, or a viral protein with internal deletion, which has the ability to form VLPs under conditions that favor VLP formation. Accordingly, the polypeptide may comprise the full-length sequence, fragments, truncated and partial sequences, as well as analogs and precursor forms of the reference molecule. The term therefore includes deletions, additions and substitutions to the sequence, so long as the polypeptide retains the ability to form a VLP. Thus, the term includes natural variations of the specified polypeptide since variations in coat proteins often occur between viral isolates. The term also includes deletions, addition and substitutions that do not naturally occur in the reference protein, so long as the protein retains the ability to form a VLP.

Preferred substitutions are those which are conservative in nature, *i.e.*, those substitutions that take place within a family of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic: aspartate and glutamate; (2) basic: lysine, arginine, and histidine; (3) non-polar: alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar: glycine, asparagine, glutamine,

cystine, serine, theronine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids. For example, it is reasonably predictable that an isolated replacement of leucine with isoleucine or valine, an asparate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid, will not have a major effect on the biological activity. Proteins having substantially the same amino acid sequence as the reference molecule, but possessing minor amino acid substitutions that do not substantially affect the immunogenicity of the protein, are therefore within the definition of the reference polypeptide.

The VLPs of the invention can be formed from any viral protein, particle-forming polypeptide derived from the viral protein, or combination of viral proteins or fragments thereof, that have the capability of forming particles under appropriate conditions. The requirements for the particle-forming viral proteins are that if the particle is formed in the cytoplasm of the host cell, the protein must be sufficiently stable in the host cell in which it is expressed such that formation of virus-like structures will result, and that the polypeptide will automatically assemble into a virus-like structure in the cell of the recombinant expression system used. If the protein is secreted into culture media, conditions can be adjusted such that VLPs will form. Furthermore, the particle-forming protein should not be cytotoxic in the expression host and should not be able to replicate in the host in which the VLP will be used.

Preferred particle-forming polypeptides include coronavirus M and E proteins, preferably SARS M and E proteins.

Methods and suitable conditions for forming particles from a wide variety of viral proteins are known in the art. VLPs have been produced, for example from proteins derived from influenza virus (such as HA or NA), Hepatitis B virus (such as core or capsid proteins), Hepatitis E virus, measles virus, Sindbis virus, Rotavirus, Foot-and-Mouth Disease virus, Retrovirus, Norwalk virus, human Papilloma virus, HIV, RNA-phages, Qß-phage (such as coat proteins), GA-phage, fr-phage, AP205 phage, and Ty (such as retrotransposon Ty protein p1). VLPs are discussed further in WO 03/024480, WO 03/024481, and Niikura *et al.*, *Virology* (2002) 293:273-280; Lenz *et al.*, *J. Immunology* (2001) 5246-5355; Pinto, *et al.*, *J. Infectious Diseases* (2003) 188:327-338; and Gerber *et al.*, *J. Virology* (2001) 75(10):4752-4760.

As explained above, VLPs can spontaneously form when the particle-forming polypeptide of interest is recombinantly expressed in an appropriate host cell. Thus, the VLPs for use in the present invention may be prepared using recombinant techniques, well known in the art. In this regard, genes encoding the particle-forming polypeptide in question can be isolated from DNA libraries or directly from cells and tissues containing the same, using known techniques. The genes encoding the particle-forming polypeptides can also be produced synthetically, based on the known sequences. The nucleotide sequence can be designed with the appropriate codons for

the particular amino sequence desired. In general, one will select preferred codons for the intended host in which the sequence will be expressed (*e.g.* human codons for human DNA vaccines). The complete sequence is generally assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See., *e.g.*, Edge, *Nature* (1981) 292:756; Nambair *et al. Science* (1984) 223:1299; Jay *et al.*, *J. Biol. Chem.* (1984) 259:6311.

Once the coding sequences for the desired particle-forming polypeptides have been isolated or synthesized, they can be cloned into any suitable vector or replicon for expression. Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. See, generally, Sambrook *et al.* The vector is then used to transform an appropriate host cell. Suitable expression systems include, but are not limited to, bacterial, mammalian, bacuolvirus/insect, vaccinia, Semliki Forest virus (SFV), yeast, and Xenopus expression systems, well known in the art.

A number of cell lines suitable for use as host cells for producing the VLPs of the invention are known in the art. Suitable mammalian cell lines include, but are not limited to, Chinese Hamster Ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*e.g.*, Hep G2), Madin-Darby bovine kidney ("MDBK") cells, as well as others. Mammalian sources of cells include, but are not limited to, human or non-human primate (*e.g.*, MRC-5 (ATCC CCL-171), WI-38 (ATCC CCL-75), HUH, human embryonic kidney cells (293 cells, typically transformed by sheared adenovirus type 5 DNA), VERO cells from monkey kidneys (including, for example COS7 cells), horse, cow (*e.g.*, MDBK cells), sheep, dog (*e.g.*, MDCK cells from dog kidneys, ATCC CCL34 MDCK (NBL2) or MDCK 33016, deposit number DSM ACC 2219 as described in WO 97/37001), cat, and rodent (*e.g.*, hamster cells such as BHK21-F, HKCC cells, or Chinese hamster ovary cells (CHO cells)), and may be obtained from a wide variety of developmental stages, including for example, adult, neonatal, fetal, and embryo.

Bacterial hosts suitable for production of VLPs of the invention include *E. coli*, *Bacillus subtilis*, and *Streptoccocus spp.* Yeast hosts suitable for production of VLPs of the invention include *Saccharomyces cerevisiae*, *Candida albicans*, *Candida maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guillerimondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica.* Insect cells suitable for production of VLPs of the invention (*i.e.*, via baculovirus expression vectors) include *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodptera frugiperda*, and *Trichoplusia ni.*

Viral vectors can be used for the production of particles in eukaryotic cells, such as those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. Additional,

vaccinia based infection/transfection systems, such as those as described in Tomei *et al.*, *J. Virol* (1993) 67:4017-4026 and Selby *et al.*, *J. Gen. Virol.* (1993) 74:1103-1113, can also be used to generate the VLPs of the invention. In this system, cells are first transfected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the DNA of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translation machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products.

Depending on the expression system and host selected, the VLPs are produced by growing host cells transformed by an expression vector under conditions whereby the particle-forming polypeptide is expressed and VLPs can be formed. The selection of the appropriate growth conditions is within the skill of the art. If the VLPs are formed intracellularly, the cells are then disrupted, using chemical, physical or mechanical means, which lyse the cells yet keep the VLPs substantially intact. Such methods are known the those of skill in the art and are described in, *e.g.*, *Protein Purification Applications: A Practical Approach*, (E.L.V. Harris and S. Angal, Eds., 1990).

The particles are then isolated using methods that preserve the integrity thereof, such as by gradient centrifugation, *e.g.*, cesium chloride (CsCl) and sucrose gradients, and the like (see, *e.g.*, Kirnbauer *et al.*, *J. Virol.* (1993) 67:6929-6936), ion exchange chromatography (including anion exchange chromatography such as DMAE and TMAE), hydroxyapatitie chromatography (see WO 00/09671), hydrophobic interaction chromatography, gel filtration chromatography and other filtration methods such as nanometric filtration and ultrafiltration. Preferably at least one anion exchange step is performed during purification, and more preferably at least two anion exchange steps are used.

VLP formulations of the invention may be further processed by methods known in the art to disassemble the VLPs into smaller, protein containing moieties using a high concentration of reducing agent, followed by reassembly of the VLPs by either removal of the reducing agent or by addition of excess oxidant. The resulting reassembled VLPs may have improved homogeneity, stability and immunogenic properties. In addition, further therapeutic or prophylactic agents may be formulated into the VLPs upon reassembly. See McCarthy *et al.*, *J. Virology* (1998) 72(1):32-41. See also WO 99/13056 and WO 01/42780. Reducing agents suitable for use in VLP disassembly include sulfhydryl reducing agents (such as glutathion, beta mercaptoethanol, dithiothreitol, dithioerythritol, cysteine, hydrogen sulfide and mixtures thereof) preferably contained in moderate to low ionic strength buffers. Sufficient exposure time of the VLPs to the reducing agent will be required to achieve a suitable amount of VLP disassembly.

Adjuvants may be added to the VLPs of the invention to enhance the immunogenicity of the SARS viral antigens. Antigens suitable for use with VLPs include those described, supra. For example, the VLPs of the invention may be adsorbed onto an aluminum adjuvant.

The VLPs of the invention may formulated to enhance their stability. Additional components which may enhance the stability of a VLP formulation include salts, buffers, non-ionic surfactants and other stabilizers such as polymeric polyanion stabilizers. See WO 00/45841.

The ionic strength of a solution comprising VLP particles may be maintained by the presence of salts. Almost any salt which can contribute to the control of the ionic strength may be used. Preferred salts which can be used to adjust ionic strength include physiologically acceptable salts such as NaCl, KCl, Na_2SO_4, $(NH_4)_2SO_4$, sodium phosphate and sodium citrate. Preferably, the salt component is present in concentrations of from about 0.10 M to 1 M. Very high concentrations are not preferred due to the practical limitations of parenteral injection of high salt concentrations. Instead, more moderate salt concentrations, such as more physiological concentrations of about 0.15M to about 0.5M with 0.15M-0.32M NaCl are preferred.

Buffers may also be used to enhance the stability of the VLP formulations of the invention. Preferably, the buffer optimizes the VLP stability while maintaining the pH range so that the vaccine formulation will not be irritating to the recipient. Buffers preferably maintain the pH of the vaccine formulation within a range of p/H 5.5-7.0, more preferably 6.0-6.5. Buffers suitable for vaccine formulations are known in the art and include, for example, histidine and imidazole. Preferably, the concentration of the buffer will range from about 2mM to about 100 mM, more preferably 5 mM to about 20 mM. Phosphate containing buffers are generally not preferred when the VLP is adsorbed or otherwise formulated with an aluminum compound.

Non-ionic surfactants may be used to enchance the stability of the VLP formulations of the invention. Surfactants suitable for use in vaccine formulations are known in the art and include, for example, polyoxyethylene sorbital fatty acid esters (Polysorbates) such as Polysorbate 80 (*e.g.*, TWEEN 80), Polysorbate 20 (*e.g.*, TWEEN 20), polyoxyethylene alkyl ethers (*e.g.*, Brij 35, Brij 58), as well as others, including Triton X-100, Triton X-114, NP-40, Span 85 and the Pluronic series of non-ionic surfactants (*e.g.*, Pluronic 121). The surfactant is preferably present in a concentration of from about 0.0005% to about 0.5% (wt/vol).

Polymeric polyanion stabilizers may also be used to enchance the stability of the VLP formulations of the invention. Suitable polymeric polyanionic stabilizers for use in the invention comprise either a single long chain or multiple cross linked chains; either type possessing multiple negative charges along the chains when in solution. Examples of suitable polyanionic polymers include proteins, polyanions, peptides and polynucelic acids. Specific examples include carboxymethyl cellulose, heparin, polyamino acids (such as poly(Glu), poly(Asp), and

Poly (Glu, Phe), oxidized glutathione, polynuceltodies, RNA, DNA and serum albumins. The concentration of the polmeric polyanion stabilizers is preferably from about 0.01% to about 0.5%, particularly about 0.05-0.1% (by weight).

G. Passive Immunization via Antibodies to the SARS Antigens of the Invention

The invention includes antibodies specific to the SARS antigens of the invention and methods of treatment or prevention of SARS virus related disease by administrating an effective amount of SARS antibodies to a mammalian subject. Antibodies specific the SARS antigens can be produced by one skilled in the art. Preferably, the antibodies are specific to the spike (S) protein of the SARS virus. Potent neutralization of the SARS coronavirus using a human monoclonal anti-spike antibody has been reported (Sui *et al.* (2004) *PNAS USA* 101:2536-2541). A IgG1 form of the monoclonal antibody showed a higher affinity (1.59 nM) than a scFv form (32.3 nM).

The antibodies of the invention are specific and selective to SARS antigens.

In one embodiment, the antibodies of the invention are generated by administering a SARS antigen to an animal. The method may also include isolating the antibodies from the animal.

The antibodies of the invention may be polyclonal or monoclonal antibody preparations, monospecific antisera, human antibodies, or may be hybrid or chimeric antibodies, such as humanized antibodies, altered antibodies $(Fab')_2$ fragments, F(ab) fragments, Fv fragments, single-domain antibodies, dimeric or trimeric antibody fragments or constructs, minibodies, or functional fragments thereof which bind to the antigen in question.

Antibodies are produced using techniques well known to those of skill in the art and disclosed in, for example, US Patent Nos. 4,011,308; 4,722,890; 4,016,043; 3,876,504; 3,770,380; and 4,372,745. For example, polyclonal antibodies are generated by immunizing a suitable animal, such as a mouse, rat, rabbit, sheep, or goat, with an antigen of interest. In order to enhance immunogenicity, the antigen can be linked to a carrier prior to immunization. Such carriers are well known to those of ordinary skill in the art. Immunization is generally performed by mixing or emulsifying the antigen in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). The animal is generally boosted 2-6 weeks later with one or more injections of the antigen in saline, preferably using Freund's incomplete adjuvant. Antibodies may also be generated by in vitro immunization, using methods known in the art. Polyclonal antiserum is then obtained from the immunized animal.

Monoclonal antibodies are generally prepared using the method of Kohler & Milstein (1975) *Nature* 256:495-497, or a modification thereof. Typically, a mouse or rat is immunized as described above. Rabbits may also be used. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into

single cells. If desired, the spleen cells may be screened (after removal of non-specifically adherent cells) by applying a cell suspension to a plate or well coated with the antigen. B-cells, expressing membrane-bound immunoglobulin specific for the antigen, will bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (*e.g.*, hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected monoclonal antibody-secreting hybridomas are then cultured either *in vitro* (*e.g.*, in tissue culture bottles or hollow fiber reactors), or *in vivo* (*e.g.*, as ascites in mice).

Humanized and chimeric antibodies are also useful in the invention. Hybrid (chimeric) antibody molecules are generally discussed in Winter *et al.* (1991) *Nature* 349: 293-299 and US Patent No. 4,816,567. Humanized antibody molecules are generally discussed in Riechmann *et al.* (1988) *Nature* 332:323-327; Verhoeyan *et al.* (1988) *Science* 239:1534-1536; and U.K. Patent Publication No. GB 2,276,169, published 21 September 1994). One approach to engineering a humanized antibody involves cloning recombinant DNA containing the promoter, leader, and variable-region sequences from a mouse antibody gene and the constant-region exons from a human antibody gene to create a mouse-human chimera, a humanized antibody. See generally, Kuby, "Immunology, 3rd Edition", W.H. Freeman and Company, New York (1998) at page 136.

Antibody fragments which retain the ability to recognize a SARS antigen are also included within the scope of the invention. A number of antibody fragments are known in the art which comprise antigen-binding sites capable of exhibiting immunological binding properties of an intact antibody molecule. For example, functional antibody fragments can be produced by cleaving a constant region, not responsible for antigen binding, from the antibody molecule, using *e.g.*, pepsin, to produce $F(ab')_2$ fragments. These fragments will contain two antigen binding sites, but lack a portion of the constant region from each of the heavy chains. Similarly, if desired, Fab fragments, comprising a single antigen binding site, can be produced, *e.g.*, by digestion of polyclonal or monclonal antibodies with papain. Functional fragments, includnig only the variable regions of the heavy and light chains, can also be produced, using standard techniques such as recombinant production or preferential proteolytic cleavage of immunoglobulin molecules. These fragments are known as F_v. See, *e.g.*, Inbar *et al.* (1972) *Proc. Nat. Acad. Sci USA* 69:2659-2662; Hochman *et al.* (1976) *Biochem* 15:2706-2710; and Ehrlich *et al.* (1980) *Biochem* 19:4091-4096.

A single-chain Fv ("sFv" or scFv") polypeptide is a covalently linked V_H-V_L heterodimer which is expressed from a gene fusion including V_H-and V_L- encoding genes linked by a peptide-

encoding linker. Huston *et al.* (1988) *Proc. Nat. Acad. Sci. USA* 85:5879-5883. A number of methods have been described to discern and develop chemical strucutres (linkers) for converting the naturally aggregated, but chemically separated, light and heavy polypeptide chains from an antibody V region into an sFv molecule which will fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. See, *e.g.*, US Patent Nos. 5,091,513; 5,132,405; and 4,946,778. The sFv molecules may be produced using methods described in thea rt. See, *e.g.*, Huston *et al.* (1988) *Proc. Nat. Acad. Sci USA* 85:5879-5338; US Patent Nos. 5,091,513; 5,132,405 and 4,946,778. Design criteria include determining the appropriate length to span the distance between the C-terminus of one chain and the N-terminus of the other, wherein the linker is generally formed from small hydrophilic amino acid residues that do not coil or form secondary structures. Such methods have been described in the art. See, *e.g.*, US Patent Nos. 5,091,513; 5,132,405 and 4,946,778. Suitable linkers generally comprise polypeptide chains of alternating sets of glycine and serine residues, and may include glutamic acid and lysine residues inserted to enhance solubility. Anti-spike scFv antibodies have been reported (Sui *et al.* (2004) *PNAS USA* 101:2536-2541).

"Mini-antibodies" or "minibodies" will also find use with the present invention. Minibodies are sFv polypeptide chains which include oligomerization domains at their C-termini, separated from the sFv by a hinge region. Pack *et al.*, (1992) *Biochem* 31:1579-1584. The oligomerization domain comprises self-associating α-helices, *e.g.*, leucine zippers, that can be further stabilized by additional disulfide bonds. The oligomerization domain is designed to be compatible with vectorial folding across a membrane, a process thought to facilitate in vivo folding of the polypeptide into a functional binding protein. Generally, minibodies are produced using recombinant methods well known in the art. See, *e.g.*, Pack *et al.*, (1992) *Biochem* 31:1579-1584; Cumber *et al.* (1992) *J. Immunology* 149B:120-126.

Non-conventional means can also be used to generate and identify the antibodies of the invention. For example, a phage display library can be screened for antibodies which bind to the SARS antigens of the invention. See generally, Siegel, "Recombinant Monoclonal Antibody Technology", *Transfus. Clin. Biol.* (2002) 9(1): 15-22; Sidhu, "Phage Display in Pharmaceutical Biotechnology", *Curr. Opin. Biotechnol.* (2000) 11(6):610-616; Sharon, *et al.*, "Recombinant Polyclonal Antibody Libraries", *Comb. Chem. High Throughput Screen* (2000) 3(3): 185-196; and Schmitz *et al.*, "Phage Display: A Molecular Tool for the Generation of Antibodies-Review", *Placenta*, (2000) 21 SupplA: S106-12.

The antibodies of the invention may also be generated by administering the polynucleotide sequence encoding for the SARS antigen into an animal. The SARS antigen is then expressed in vivo, and antibodies specific to the SARS antigen are generated *in vivo*. Methods for polynucleotide delivery of the SARS antigens of the invention are discussed in section 4 below.

The antibodies of the invention are preferably specific to the SARS virus.

H. Combinations of one or more of any of the above approaches in a vaccine

The compositions of the invention further comprise combinations of one or more of the compositions discussed above. For instance, the invention comprises a composition comprising an attenuated SARS virus and a subunit SARS viral antigen.

I. Combinations of SARS antigens and other Respiratory Virus Antigens

The invention further relates to vaccine formulations comprising one or more SARS virus antigens and one or more other respiratory virus antigens. Additional respiratory virus antigens suitable for use in the invention include antigens from influenza virus, human rhinovirus (HRV), parainfluenza virus (PIV), respiratory syncytial virus (RSV), adenovirus, metapneumovirus, and rhinovirus. The additional respiratory virus antigen could also be from a coronavirus other than the SARS coronavirus, such as the NL63 human coronavirus (van der Hoek *et al.* (2004) *Nature Medicine* 10:368-373). Preferably, the additional respiratory virus antigen is an influenza viral antigen.

The invention may also comprise one or more bacterial or viral antigens in combination with the SARS viral antigen. Antigens may be used alone or in any combination. (See, *e.g.*, WO 02/00249 describing the use of combinations of bacterial antigens). The combinations may include multiple antigens from the same pathogen, multiple antigens from different pathogens or multiple antigens from the same and from different pathogens. Thus, bacterial, viral, and/or other antigens may be included in the same composition or may be administered to the same subject separately. It is generally preferred that combinations of antigens be used to raise an immune response be used in combinations.

Non-limiting examples of bacterial pathogens which may be used in the invention include diphtheria (*See, e.g.,* Chapter 3 of *Vaccines*, 1998, eds. Plotkin & Mortimer (ISBN 0-7216-1946-0), staphylococcus (*e.g., Staphylococcus aureus* as described in Kuroda *et al.* (2001) *Lancet* 357:1225-1240), cholera, tuberculosis, *C. tetani,* also known as tetanus (*See, e.g.,* Chapter 4 of *Vaccines*, 1998, eds. Plotkin & Mortimer (ISBN 0-7216-1946-0), Group A and Group B streptococcus (including *Streptococcus pneumoniae, Streptococcus agalactiae* and *Streptococcus pyogenes* as described, for example, in Watson *et al.* (2000) *Pediatr. Infect. Dis. J.* 19:331-332; Rubin *et al.* (2000) *Pediatr Clin. North Am.* 47:269-284; Jedrzejas *et al.* (2001) *Microbiol Mol Biol Rev* 65:187-207; Schuchat (1999) *Lancet* 353:51-56; GB patent applications 0026333.5; 0028727.6; 015640.7; Dale *et al.* (1999) *Infect Dis Clin North Am* 13:227-1243; Ferretti *et al.* (2001) *PNAS USA* 98:4658-4663), pertussis (*See, e.g.,* Gusttafsson *et al.* (1996) *N. Engl. J. Med.* 334:349-355; Rappuoli *et al.* (1991) *TIBTECH* 9:232-238), meningitis, *Moraxella catarrhalis* (*See, e.g.,* McMichael (2000) *Vaccine* 19 Suppl. 1:S101-107) and other pathogenic states,

including, without limitation, Neisseria meningitides (A, B, C, Y), *Neisseria gonorrhoeae* (*See, e.g.,* WO 99/24578; WO 99/36544; and WO 99/57280), Helicobacter pylori (*e.g.,* CagA, VacA, NAP, HopX, HopY and/or urease as described, for example, WO 93/18150; WO 99/53310; WO 98/04702) and *Haemophilus influenza.* Hemophilus influenza type B (HIB) (See, *e.g.,* Costantino *et al.* (1999) *Vaccine* 17:1251-1263), *Porphyromonas gingivalis* (Ross *et al.* (2001) *Vaccine* 19:4135-4132) and combinations thereof.

Non-limiting examples of viral pathogens which may be used in the invention include meningitis, rhinovirus, influenza (Kawaoka *et al.*, *Virology* (1990) 179:759-767; Webster *et al.*, "Antigenic variation among type A influenza viruses," p. 127-168. In: P. Palese and D.W. Kingsbury (ed.), Genetics of influenza viruses. Springer-Verlag, New York), respiratory syncytial virus (RSV), parainfluenza virus (PIV), rotavirus (*e.g.,* VP1, VP2, VP3, VP4, VP6, VP7, NSP1, NSP2, NSP3, NSP4 or NSP5 and other rotavirus antigens, for example as described in WO 00/26380) and the like. Antigens derived from other viruses will also find use in the present invention, such as without limitation, proteins from members of the families Picomaviridae (*e.g.*, polioviruses, *etc.* as described, for example, in Sutter *et al.* (2000) *Pediatr Clin North Am* 47:287-308; Zimmerman & Spann (1999) *Am Fam Physician* 59:113-118; 125-126); Caliciviridae; Togaviridae (*e.g.*, rubella virus, *etc.*); Flaviviridae, including the genera flavivirus (*e.g.,* yellow fever virus, Japanese encephalitis virus, serotypes of Dengue virus, tick borne encephalitis virus, West Nile virus, St. Louis encephalitis virus); pestivirus (*e.g.,* classical porcine fever virus, bovine viral diarrhea virus, border disease virus); and hepacivirus (*e.g.,* hepatitis A, B and C as described, for example, in US Patent Nos. 4,702,909; 5,011,915; 5,698,390; 6,027,729; and 6,297,048); Parvovirus (*e.g.,* parvovirus B19); Coronaviridae; Reoviridae; Bimaviridae; Rhabodoviridae (*e.g.*, rabies virus, *etc.* as described for example in Dressen *et al.* (1997) *Vaccine* 15 Suppl:s2-6; MMWR Morb Mortal Wkly Rep. 1998 Jan 16:47(1):12, 19); Filoviridae; Paramyxoviridae (*e.g.*, mumps virus, measles virus, respiratory syncytial virus, *etc.* as described in Chapters 9 to 11 of *Vaccines*, 1998, eds. Plotkin & Mortimer (ISBN 0-7216-1946-0); Orthomyxoviridae (*e.g.*, influenza virus types A, B and C, *etc.* as described in Chapter 19 of *Vaccines*, 1998, eds. Plotkin & Mortimer (ISBN 0-7216-1946-0),.); Bunyaviridae; Arenaviridae; Retroviradae (*e.g.*, HTLV-1; HTLV-11; HIV-1 (also known as HTLV-III, LAV, ARV, HTI,R, *etc.*)), including but not limited to antigens from the isolates HIVIIIb, HIVSF2, HIVLAV, HIVI-AL, I-IIVMN, SF162); HIV- I CM235, HIV- I US4; HIV-2; simian immunodeficiency virus (SIV) among others. Additionally, antigens may also be derived from human papilloma virus (HPV) and the tick-borne encephalitis viruses. See, *e.g.* Virology, 3rd Edition (W.K. Joklik ed. 1988); Fundamental Virology, 2nd Edition (B.N. Fields and D.M. Knipe, eds, 1991), for a description of these and other viruses.

Proteins may also be derived from the herpesvirus family, including proteins derived from herpes simplex virus (HSV) types 1 and 2, such as HSV-1 and HSV-2 glycoproteins gB, gD and gH; antigens derived from varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus (CMV) including CMV gB and gH (See, US Patent No. 4,689,225 and PCT Publication WO 89/07143); and antigens derived from other human herpesviruses such as HHV6 and HHV7. (See, *e.g.* Chee *et al.*, *Cytomegaloviruses* (J.K. McDougall, ed., Springer-Verlag 1990) pp. 125-169, for a review of the protein coding content of cytomegalovirus; McGeoch *et al.*, *J. Gen. Virol.* (1988) 69:1531-1574, for a discussion of the various HSV-1 encoded proteins; US Patent No. 5,171,568 for a discussion of HSV-1 and HSV-2 gB and gD proteins and the genes encoding therefor; Baer *et al.*, *Nature* (1984) 310:207-211, for the identification of protein coding sequences in an EBV genome; and Davison and Scott, *J. Gen. Virol.* (1986) 67:1759-1816, for a review of VZV). Herpes simplex virus (HSV) rgD2 is a recombinant protein produced in genetically engineered Chinese hamster ovary cells. This protein has the normal anchor region truncated, resulting in a glycosylated protein secreted into tissue culture medium. The gD2 can be purified in the CHO medium to greater than 90% purity. Human immunodeficiency virus (HIV) env-2-3 is a recombinant form of the HIV enveloped protein produced in genetically engineered *Saccharomyces cerevisae*. This protein represents the entire protein region of HIV gp120 but is non-glycosylated and denatured as purified from the yeast. HIV gpl20 is a fully glycosylated, secreted form of gp120 produced in CHO cells in a fashion similar to the gD2 above. Additional HSV antigens suitable for use in immunogenic compositions are described in PCT Publications W0 85/04587 and W0 88/02634, the disclosures of which are incorporated herein by reference in their entirety. Mixtures of gB and gD antigens, which are truncated surface antigens lacking the anchor regions, are particularly preferred.

Antigens from the hepatitis family of viruses, including hepatitis A virus (HAV) (*See, e.g.,* Bell *et al.* (2000) *Pediatr Infect Dis. J.* 19:1187-1188; Iwarson (1995) *APMIS* 103:321-326), hepatitis B virus (HBV) (*See, e.g.,* Gerlich *et al.* (1990) *Vaccine* 8 Suppl:S63-68 & 79-80), hepatitis C virus (HCV) (*See, e.g.,* PCT/US88/04125, published European application number 318216), the delta hepatitis virus (HDV), hepatitis E virus (HEV) and hepatitis G virus (HGV), can also be conveniently used in the techniques described herein. By way of example, the viral genomic sequence of HCV is known, as are methods for obtaining the sequence. See, *e.g.*, International Publication Nos. WO 89/04669; WO 90/11089; and WO 90/14436. Also included in the invention are molecular variants of such polypeptides, for example as described in PCT/US99/31245; PCT/US99/31273 and PCT/US99/31272. The HCV genome encodes several viral proteins, including E1 (also known as E) and E2 (also known as E2/NSI) and an N-terminal nucleocapsid protein (termed "core") (see, Houghton *et al.*, *Hepatology* (1991) 14:381-388, for a discussion of HCV proteins, including E1 and E2). Similarly, the sequence for the δ-antigen

from HDV is known (see, *e.g.*, US Patent No. 5,378,814) and this antigen can also be conveniently used in the present composition and methods. Additionally, antigens derived from HBV, such as the core antigen, the surface antigen, SAg, as well as the presurface sequences, pre-S1 and pre-S2 (formerly called pre-S), as well as combinations of the above, such as SAg/pre-S1, SAg/pre-S2, SAg/pre-S1/pre-S2, and pre-S1/pre-S2, will find use herein. See, *e.g.*, "HBV Vaccines - from the laboratory to license: a case study" in Mackett, M. and Williamson, J.D., *Human Vaccines and Vaccination*, pp. 159-176, for a discussion of HBV structure; and US Patent Nos. 4,722,840, 5,098,704, 5,324,513, incorporated herein by reference in their entireties; Beames *et al.*, *J. Virol.* (1995) 69:6833-6838, Birnbaum *et al.*, *J. Virol.* (1990) 64:3319-3330; and Zhou *et al.*, *J. Virol.* (1991) 65:5457-5464. Each of these proteins, as well as antigenic fragments thereof, will find use in the present composition and methods.

Influenza virus is another example of a virus for which the present invention will be particularly useful. Specifically, the envelope glycoproteins HA and NA of influenza A are of particular interest for generating an immune response. Numerous HA subtypes of influenza A have been identified (Kawaoka *et al.*, *Virology* (1990) 179:759-767; Webster *et al.*, "Antigenic variation among type A influenza viruses," p. 127-168. In: P. Palese and D.W. Kingsbury (ed.), *Genetics of influenza viruses*. Springer-Verlag, New York). Thus, proteins derived from any of these isolates can also be used in the compositions and methods described herein.

Non-limiting examples of parasitic antigens include those derived from organisms causing malaria and Lyme disease.

The methods of the invention comprise administering an immunogenic composition comprising a SARS viral antigen (including one or more of an inactivated SARS virus, an attenuated SARS virus, a split SARS virus preparation or a recombinant or purified subunit formulation of one or more SARS viral antigens) to an animal. The immunogenic compositions used in the invention can comprise an immunologically effective amount of the SARS viral antigen. An "immunologically effective amount" is an amount sufficient to allow the mammal to raise an immune response to the SARS antigen.

The immune response preferably involves the production of antibodies specific to the SARS antigen. The amount of antibodies produced will vary depending on several factors including the animal used, the presence of an adjuvant, *etc.*

The immunogenic compositions of the invention may further comprise one or more adjuvants.

The immunogenic compositions of the invention may be administered mucosally. Suitable routes of mucosal administration include oral, intranasal, intragastric, pulmonary, intestinal, rectal, ocular and vaginal routes. The immunogenic composition may be adapted for mucosal administration. For instance, where the composition is for oral administration, it may be in the

form of tablets or capsules, optionally enteric-coated, liquid, transgenic plants, *etc.* Where the composition is for intranasal administration, it may be in the form of a nasal spray, nasal drops, gel or powder.

The immunogenic compositions of the invention may be administered parenterally. Suitable routes of parenteral administration include intramuscular (IM), subcutaneous, intravenous, intraperitoneal, intradermal, transcutaneous, and transdermal (*see e.g.*, International patent application WO 98/20734) routes, as well as delivery to the interstitial space of a tissue. The immunogenic composition may be adapted for parenteral administration, for instance in the form of an injectable that may be sterile and pyrogen free.

Vaccines of the invention may be administered in conjunction with other immunoregulatory agents. In particular, compositions will usually include an adjuvant. Preferred further adjuvants include, but are not limited to, one or more of the following set forth below:

A. Mineral Containing Compositions

Mineral containing compositions suitable for use as adjuvants in the invention include mineral salts, such as aluminium salts and calcium salts. The invention includes mineral salts such as hydroxides (*e.g.* oxyhydroxides), phosphates (*e.g.* hydroxyphoshpates, orthophosphates), sulphates, *etc.* (*e.g.* see chapters 8 & 9 of *Vaccine design:the subunit and adjuvant approach* (1995) Powell & Newman. ISBN 0-306-44867-X.), or mixtures of different mineral compounds, with the compounds taking any suitable form (*e.g.* gel, crystalline, amorphous, *etc.*), and with adsorption being preferred. The mineral containing compositions may also be formulated as a particle of metal salt. See WO00/23105.

B. Oil-Emulsions

Oil-emulsion compositions suitable for use as adjuvants in the invention include squalene-water emulsions, such as MF59 (5% Squalene, 0.5% Tween 80, and 0.5% Span 85, formulated into submicron particles using a microfluidizer). See WO90/14837. See also, Frey *et al.*, "Comparison of the safety, tolerability, and immunogenicity of a MF59-adjuvanted influenza vaccine and a non-adjuvanted influenza vaccine in non-elderly adults", *Vaccine* (2003) 21:4234-4237.

Particularly preferred adjuvants for use in the compositions are submicron oil-inwater emulsions. Preferred submicron oil-in-water emulsions for use herein are squalene/water emulsions optionally containing varying amounts of MTP-PE, such as a submicron oil-in-water emulsion containing 4-5% w/v squalene, 0.25-1.0% w/v Tween 80 ™ (polyoxyelthylenesorbitan monooleate), and/or 0.25-1.0% Span 85™ (sorbitan trioleate), and, optionally, N-acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-huydroxyphosphophoryloxy)-ethylamine (MTP-PE), for example, the submicron oil-in-water

emulsion known as "MF59" (International Publication No. WO 90/14837; US Patent Nos. 6,299,884 and 6,451,325, incorporated herein by reference in their entireties; and Ott *et al.*, "MF59 -- Design and Evaluation of a Safe and Potent Adjuvant for Human Vaccines" in *Vaccine Design: The Subunit and Adjuvant Approach* (Powell, M.F. and Newman, M.J. eds.) Plenum Press, New York, 1995, pp. 277-296). MF59 contains 4-5% w/v Squalene (*e.g.*, 4.3%), 0.25-0.5% w/v Tween 80™, and 0.5% w/v Span 85™ and optionally contains various amounts of MTP-PE, formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA). For example, MTP-PE may be present in an amount of about 0-500 μg/dose, more preferably 0-250 μg/dose and most preferably, 0-100 μg/dose. As used herein, the term "MF59-0" refers to the above submicron oil-in-water emulsion lacking MTP-PE, while the term MF59-MTP denotes a formulation that contains MTP-PE. For instance, "MF59-100" contains 100 μg MTP-PE per dose, and so on. MF69, another submicron oil-in-water emulsion for use herein, contains 4.3% w/v squalene, 0.25% w/v Tween 80™, and 0.75% w/v Span 85™ and optionally MTP-PE. Yet another submicron oil-in-water emulsion is MF75, also known as SAF, containing 10% squalene, 0.4% Tween 80™, 5% pluronic-blocked polymer L121, and thr-MDP, also microfluidized into a submicron emulsion. MF75-MTP denotes an MF75 formulation that includes MTP, such as from 100-400 μg MTP-PE per dose.

Submicron oil-in-water emulsions, methods of making the same and immunostimulating agents, such as muramyl peptides, for use in the compositions, are described in detail in International Publication No. WO 90114837 and US Patent Nos. 6,299,884 and 6,45 1,325, incorporated herein by reference in their entireties.

Complete Freund's adjuvant (CFA) and incomplete Freund's adjuvant (IFA) may also be used as adjuvants in the invention.

C. Saponin Formulations

Saponin formulations, may also be used as adjuvants in the invention. Saponins are a heterologous group of sterol glycosides and triterpenoid glycosides that are found in the bark, leaves, stems, roots and even flowers of a wide range of plant species. Saponin from the bark of the *Quillaia saponaria* Molina tree have been widely studied as adjuvants. Saponin can also be commercially obtained from *Smilax ornata* (sarsaprilla), *Gypsophilla paniculata* (brides veil), and *Saponaria officianalis* (soap root). Saponin adjuvant formulations include purified formulations, such as QS21, as well as lipid formulations, such as ISCOMs.

Saponin compositions have been purified using High Performance Thin Layer Chromatography (HP-LC) and Reversed Phase High Performance Liquid Chromatography (RP-HPLC). Specific purified fractions using these techniques have been identified, including QS7,

QS17, QS18, QS21, QH-A, QH-B and QH-C. Preferably, the saponin is QS21. A method of production of QS21 is disclosed in US Patent No. 5,057,540. Saponin formulations may also comprise a sterol, such as cholesterol (see WO 96/33739).

Combinations of saponins and cholesterols can be used to form unique particles called Immunostimulating Complexs (ISCOMs). ISCOMs typically also include a phospholipid such as phosphatidylethanolamine or phosphatidylcholine. Any known saponin can be used in ISCOMs. Preferably, the ISCOM includes one or more of Quil A, QHA and QHC. ISCOMs are further described in EP 0 109 942, WO 96/11711 and WO 96/33739. Optionally, the ISCOMS may be devoid of additional detergent. See WO00/07621.

A review of the development of saponin based adjuvants can be found at Barr, *et al.*, "ISCOMs and other saponin based adjuvants", Advanced Drug Delivery Reviews (1998) 32:247-271. See also Sjolander, *et al.*, "Uptake and adjuvant activity of orally delivered saponin and ISCOM vaccines", Advanced Drug Delivery Reviews (1998) 32:321-338.

D. Bacterial or Microbial Derivatives

Adjuvants suitable for use in the invention include bacterial or microbial derivatives such as:

(1) Non-toxic derivatives of enterobacterial lipopolysaccharide (LPS)

Such derivatives include Monophosphoryl lipid A (MPL) and 3-O-deacylated MPL (3dMPL). 3dMPL is a mixture of 3 De-O-acylated monophosphoryl lipid A with 4, 5 or 6 acylated chains. A preferred "small particle" form of 3 De-O-acylated monophosphoryl lipid A is disclosed in EP 0 689 454. Such "small particles" of 3dMPL are small enough to be sterile filtered through a 0.22 micron membrane (see EP 0 689 454). Other non-toxic LPS derivatives include monophosphoryl lipid A mimics, such as aminoalkyl glucosaminide phosphate derivatives *e.g.* RC-529. See Johnson *et al.* (1999) *Bioorg Med Chem Lett* 9:2273-2278.

(2) Lipid A Derivatives

Lipid A derivatives include derivatives of lipid A from *Escherichia coli* such as OM-174. OM-174 is described for example in Meraldi *et al.*, "OM-174, a New Adjuvant with a Potential for Human Use, Induces a Protective Response with Administered with the Synthetic C-Terminal Fragment 242-310 from the circumsporozoite protein of Plasmodium berghei", Vaccine (2003) 21:2485-2491; and Pajak, *et al.*, "The Adjuvant OM-174 induces both the migration and maturation of murine dendritic cells in vivo", Vaccine (2003) 21:836-842.

(3) Immunostimulatory oligonucleotides

Immunostimulatory oligonucleotides suitable for use as adjuvants in the invention include nucleotide sequences containing a CpG motif (a sequence containing an unmethylated cytosine followed by guanosine and linked by a phosphate bond). Bacterial double stranded RNA or

oligonucleotides containing palindromic or poly(dG) sequences have also been shown to be immunostimulatory.

The CpG's can include nucleotide modifications/analogs such as phosphorothioate modifications and can be double-stranded or single-stranded. Optionally, the guanosine may be replaced with an analog such as 2'-deoxy-7-deazaguanosine. See Kandimalla, *et al.*, "Divergent synthetic nucleotide motif recognition pattern: design and development of potent immunomodulatory oligodeoxyribonucleotide agents with distinct cytokine induction profiles", Nucleic Acids Research (2003) 31(9): 2393-2400; WO 02/26757 and WO 99/62923 for examples of possible analog substitutions. The adjuvant effect of CpG oligonucleotides is further discussed in Krieg, "CpG motifs: the active ingredient in bacterial extracts?", Nature Medicine (2003) 9(7): 831-835; McCluskie, *et al.*, "Parenteral and mucosal prime-boost immunization strategies in mice with hepatitis B surface antigen and CpG DNA", FEMS Immunology and Medical Microbiology (2002) 32:179-185; WO 98/40100; US Patent No. 6,207,646; US Patent No. 6,239,116 and US Patent No. 6,429,199.

The CpG sequence may be directed to TLR9, such as the motif GTCGTT or TTCGTT. See Kandimalla, *et al.*, "Toll-like receptor 9: modulation of recognition and cytokine induction by novel synthetic CpG DNAs", Biochemical Society Transactions (2003) 31 (part 3): 654-658. The CpG sequence may be specific for inducing a Th1 immune response, such as a CpG-A ODN, or it may be more specific for inducing a B cell response, such a CpG-B ODN. CpG-A and CpG-B ODNs are discussed in Blackwell, *et al.*, "CpG-A-Induced Monocyte IFN-gamma-Inducible Protein-10 Production is Regulated by Plasmacytoid Dendritic Cell Derived IFN-alpha", J. Immunol. (2003) 170(8):4061-4068; Krieg, "From A to Z on CpG", TRENDS in Immunology (2002) 23(2): 64-65 and WO 01/95935. Preferably, the CpG is a CpG-A ODN.

Preferably, the CpG oligonucleotide is constructed so that the 5' end is accessible for receptor recognition. Optionally, two CpG oligonucleotide sequences may be attached at their 3' ends to form "immunomers". See, for example, Kandimalla, *et al.*, "Secondary structures in CpG oligonucleotides affect immunostimulatory activity", BBRC (2003) 306:948-953; Kandimalla, *et al.*, "Toll-like receptor 9: modulation of recognition and cytokine induction by novel synthetic GpG DNAs", Biochemical Society Transactions (2003) 31(part 3):664-658; Bhagat *et al.*, "CpG penta- and hexadeoxyribonucleotides as potent immunomodulatory agents" BBRC (2003) 300:853-861 and WO 03/035836.

(4) ADP-ribosylating toxins and detoxified derivatives thereof.

Bacterial ADP-ribosylating toxins and detoxified derivatives thereof may be used as adjuvants in the invention. Preferably, the protein is derived from *E. coli* (*i.e.*, E. coli heat labile enterotoxin "LT), cholera ("CT"), or pertussis ("PT"). The use of detoxified ADP-ribosylating toxins as mucosal adjuvants is described in WO 95/17211 and as parenteral adjuvants in WO

98/42375. Preferably, the adjuvant is a detoxified LT mutant such as LT-K63, LT-R72, and LTR192G. The use of ADP-ribosylating toxins and detoxified derivaties thereof, particularly LT-K63 and LT-R72, as adjuvants can be found in the following references, each of which is specifically incorporated by reference herein in their entirety: Beignon, *et al.*, "The LTR72 Mutant of Heat-Labile Enterotoxin of Escherichia coli Enahnces the Ability of Peptide Antigens to Elicit CD4+ T Cells and Secrete Gamma Interferon after Coapplication onto Bare Skin", Infection and Immunity (2002) 70(6):3012-3019; Pizza, *et al.*, "Mucosal vaccines: non-toxic derivatives of LT and CT as mucosal adjuvants", Vaccine (2001) 19:2534-2541; Pizza, *et al.*, "LTK63 and LTR72, two mucosal adjuvants ready for clinical trials" Int. J. Med. Microbiol (2000) 290(4-5):455-461; Scharton-Kersten *et al.*, "Transcutaneous Immunization with Bacterial ADP-Ribosylating Exotoxins, Subunits and Unrelated Adjuvants", Infection and Immunity (2000) 68(9):5306-5313; Ryan *et al.*, "Mutants of Escherichia coli Heat-Labile Toxin Act as Effective Mucosal Adjuvants for Nasal Delivery of an Acellular Pertussis Vaccine: Differential Effects of the Nontoxic AB Complex and Enzyme Activity on Th1 and Th2 Cells" Infection and Immunity (1999) 67(12):6270-6280; Partidos *et al.*, "Heat-labile enterotoxin of Escherichia coli and its site-directed mutant LTK63 enhance the proliferative and cytotoxic T-cell responses to intranasally co-immunized synthetic peptides", Immunol. Lett. (1999) 67(3):209-216; Peppoloni *et al.*, "Mutants of the Escherichia coli heat-labile enterotoxin as safe and strong adjuvants for intranasal delivery of vaccines", Vaccines (2003) 2(2):285-293; and Pine *et al.*, (2002) "Intranasal immunization with influenza vaccine and a detoxified mutant of heat labile enterotoxin from Escherichia coli (LTK63)" J. Control Release (2002) 85(1-3):263-270. Numerical reference for amino acid substitutions is preferably based on the alignments of the A and B subunits of ADP-ribosylating toxins set forth in Domenighini *et al.*, Mol. Microbiol (1995) 15(6):1165-1167, specifically incorporated herein by reference in its entirety.

E. Human Immunomodulators

Human immunomodulators suitable for use as adjuvants in the invention include cytokines, such as interleukins (*e.g.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, *etc.*), interferons (*e.g.* interferon-γ), macrophage colony stimulating factor, and tumor necrosis factor.

F. Bioadhesives and Mucoadhesives

Bioadhesives and mucoadhesives may also be used as adjuvants in the invention. Suitable bioadhesives include esterified hyaluronic acid microspheres (Singh *et al.* (2001) *J. Cont. Rele.* 70:267-276) or mucoadhesives such as cross-linked derivatives of poly(acrylic acid), polyvinyl alcohol, polyvinyl pyrollidone, polysaccharides and carboxymethylcellulose. Chitosan and derivatives thereof may also be used as adjuvants in the invention. *E.g.*, WO99/27960.

G. Microparticles

Microparticles may also be used as adjuvants in the invention. Microparticles (*i.e.* a particle of ~100nm to ~150μm in diameter, more preferably ~200nm to ~30μm in diameter, and most preferably ~500nm to ~10μm in diameter) formed from materials that are biodegradable and non-toxic (*e.g.* a poly(α-hydroxy acid), a polyhydroxybutyric acid, a polyorthoester, a polyanhydride, a polycaprolactone, *etc.*), with poly(lactide-co-glycolide) are preferred, optionally treated to have a negatively-charged surface (*e.g.* with SDS) or a positively-charged surface (*e.g.* with a cationic detergent, such as CTAB).

H. Liposomes

Examples of liposome formulations suitable for use as adjuvants are described in US Patent No. 6,090,406, US Patent No. 5,916,588, and EP 0 626 169.

I. Polyoxyethylene ether and Polyoxyethylene Ester Formulations

Adjuvants suitable for use in the invention include polyoxyethylene ethers and polyoxyethylene esters. WO99/52549. Such formulations further include polyoxyethylene sorbitan ester surfactants in combination with an octoxynol (WO01/21207) as well as polyoxyethylene alkyl ethers or ester surfactants in combination with at least one additional non-ionic surfactant such as an octoxynol (WO01/21152).

Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether (laureth 9), polyoxyethylene-9-steoryl ether, polyoxytheylene-8-steoryl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether.

J. Polyphosphazene (PCPP)

PCPP formulations are described, for example, in Andrianov *et al.*, "Preparation of hydrogel microspheres by coacervation of aqueous polyphophazene solutions", Biomaterials (1998) 19(1-3):109-115 and Payne *et al.*, "Protein Release from Polyphosphazene Matrices", Adv. Drug. Delivery Review (1998) 31(3):185-196.

K. Muramyl peptides

Examples of muramyl peptides suitable for use as adjuvants in the invention include N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), and N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine MTP-PE).

L. Imidazoquinolone Compounds.

Examples of imidazoquinolone compounds suitable for use adjuvants in the invention include Imiquamod and its homologues, described further in Stanley, "Imiquimod and the

imidazoquinolones: mechanism of action and therapeutic potential" Clin Exp Dermatol (2002) 27(7):571-577 and Jones, "Resiquimod 3M", Curr Opin Investig Drugs (2003) 4(2):214-218.

M. Virosomes and Virus Like Particles (VLPs)

Virosomes and Virus Like Particles (VLPs) can also be used as adjuvants in the invention. These structures generally contain one or more proteins from a virus optionally combined or formulated with a phospholipid. They are generally non-pathogenic, non-replicating and generally do not contain any of the native viral genome. The viral proteins may be recombinantly produced or isolated from whole viruses. These viral proteins suitable for use in virosomes or VLPs include proteins derived from influenza virus (such as HA or NA), Hepatitis B virus (such as core or capsid proteins), Hepatitis E virus, measles virus, Sindbis virus, Rotavirus, Foot-and-Mouth Disease virus, Retrovirus, Norwalk virus, human Papilloma virus, HIV, RNA-phages, Qß-phage (such as coat proteins), GA-phage, fr-phage, AP205 phage, and Ty (such as retrotransposon Ty protein p1). VLPs are discussed further in WO 03/024480, WO 03/024481, and Niikura *et al.*, "Chimeric Recombinant Hepatitis E Virus-Like Particles as an Oral Vaccine Vehicle Presenting Foreign Epitopes", Virology (2002) 293:273-280; Lenz *et al.*, "Papillomarivurs-Like Particles Induce Acute Activation of Dendritic Cells", Journal of Immunology (2001) 5246-5355; Pinto, *et al.*, "Cellular Immune Responses to Human Papillomavirus (HPV)-16 L1 Healthy Volunteers Immunized with Recombinant HPV-16 L1 Virus-Like Particles", *Journal of Infectious Diseases* (2003) 188:327-338; and Gerber *et al.*, "Human Papillomavrisu Virus-Like Particles Are Efficient Oral Immunogens when Coadministered with Escherichia coli Heat-Labile Entertoxin Mutant R192G or CpG", Journal of Virology (2001) 75(10):4752-4760. Virosomes are discussed further in, for example, Gluck *et al.*, "New Technology Platforms in the Development of Vaccines for the Future", Vaccine (2002) 20:B10 –B16.

The invention may also comprise combinations of aspects of one or more of the adjuvants identified above. For example, the following adjuvant compositions may be used in the invention:

(1) a saponin and an oil-in-water emulsion (WO99/11241);

(2) a saponin (*e.g..*, QS21) + a non-toxic LPS derivative (*e.g.*, 3dMPL) (see WO 94/00153);

(3) a saponin (*e.g..*, QS21) + a non-toxic LPS derivative (*e.g.*, 3dMPL) + a cholesterol;

(4) a saponin (*e.g.* QS21) + 3dMPL + IL-12 (optionally + a sterol) (WO98/57659);

(5) combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions (See European patent applications 0835318, 0735898 and 0761231);

(6) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-block polymer L121, and thr-MDP, either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion.

(7) Ribi™ adjuvant system (RAS), (Ribi Immunochem) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); and

(8) one or more mineral salts (such as an aluminum salt) + a non-toxic derivative of LPS (such as 3dPML).

Aluminium salts and MF59 are preferred adjuvants for parenteral immunisation. Mutant bacterial toxins are preferred mucosal adjuvants.

As mentioned above, adjuvants suitable for use in the invention may also include one or more of the following:

- *E.coli* heat-labile enterotoxin ("LT"), or detoxified mutants thereof, such as the K63 or R72 mutants;

- cholera toxin ("CT"), or detoxified mutants thereof;

- microparticles (i.e., a particle of ~100nm to ~150μm in diameter, more preferably ~200nm to ~30μm in diameter, and most preferably ~500nm to ~10μm in diameter) formed from materials that are biodegradable and non-toxic (e.g. a poly(α-hydroxy acid), a polyhydroxybutyric acid, a polyorthoester, a polyanhydride, a polycaprolactone *etc.*);

- a polyoxyethylene ether or a polyoxyethylene ester (see International patent application WO 99/52549);

- a polyoxyethylene sorbitan ester surfactant in combination with an octoxynol (see International patent application WO 01/21207) or a polyoxyethylene alkyl ether or ester surfactant in combination with at least one additional non-ionic surfactant such as an octoxynol (see International patent application WO 01/21152);

- chitosan (e.g. International patent application WO 99/27960)

- an immunostimulatory oligonucleotide (e.g. a CpG oligonucleotide) and a saponin (see International patent application WO 00/62800)

- immunostimulatory double stranded RNA.

- aluminum compounds (e.g. aluminum hydroxide, aluminum phosphate, aluminum hydroxyphosphate, oxyhydroxide, orthophosphate, sulfate *etc.* (*e.g.* see chapters 8 & 9 of *Vaccine design: the subunit and adjuvant aproach,* eds. Powell & Newman, Plenum Press 1995 (ISBN 0-306-44867-X) (hereinafter "Vaccine design"), or mixtures of different aluminum compounds, with the compounds taking any suitable form (e.g. gel, crystalline, amorphous *etc.*), and with adsorption being preferred;

- MF59 (5% Squalene, 0.5% Tween 80, and 0.5% Span 85, formulated into submicron particles using a microfluidizer) (see Chapter 10 of Vaccine design; see also International patent application WO 90/14837);

- liposomes (see Chapters 13 and 14 of *Vaccine design*);

- ISCOMs (see Chapter 23 of *Vaccine design*);

- SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-block polymer L121, and thr-MDP, either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion (see Chapter 12 of *Vaccine design*);

- Ribi™ adjuvant system (RAS), (Ribi Immunochem) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™);

- saponin adjuvants, such as QuilA or QS21 (see Chapter 22 of *Vaccine design*), also known as Stimulon™;

- ISCOMs, which may be devoid of additional detergent (WO 00/07621);

- complete Freund's adjuvant (CFA) and incomplete Freund's adjuvant (IFA);

- cytokines, such as interleukins (e.g. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (e.g. interferon-γ), macrophage colony stimulating factor, tumor necrosis factor, etc. (see Chapters 27 & 28 of *Vaccine design*);

- monophosphoryl lipid A (MPL) or 3-O-deacylated MPL (3dMPL) (e.g. chapter 21 of *Vaccine design*);

- combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions (European patent applications 0835318, 0735898 and 0761231);

- oligonucleotides comprising CpG motifs (see Krieg (2000) Vaccine, 19:618-622; Krieg (2001) Curr. Opin. Mol. Ther., 2001, 3:15-24; WO 96/02555, WO 98/16247, WO 98/18810, WO 98/40100, WO 98/55495, WO 98/37919 and WO 98/52581, etc.) *i.e.* containing at least one CG dinucleotide,

- a polyoxyethylene ether or a polyoxyethylene ester (International patent application WO99/52549);

- a polyoxyethylene sorbitan ester surfactant in combination with an octoxynol (International patent application WO 01/21207) or a polyoxyethylene alkyl ether or ester surfactant in combination with at least one additional non-ionic surfactant such as an octoxynol (WO 01/21152);

- an immunostimulatory oligonucleotide (e.g. a CpG oligonucleotide) and a saponin (WO00/62800);

- an immunostimulant and a particle of metal salt (International patent application WO00/23105);

- a saponin and an oil-in-water emulsion (WO 99/11241);

- a saponin (e.g. QS21) + 3dMPL + IL-12 (optionally + a sterol) (WO 98/57659).

Other adjuvants suitable for mucosal or parenteral administration are also available (*e.g.* see chapter 7 of *Vaccine design: the subunit and adjuvant aproach*, eds. Powell & Newman, Plenum Press 1995 (ISBN 0-306-44867-X).

Mutants of LT are preferred mucosal adjuvants, in particular the "K63" and "R72" mutants (*e.g.* see International patent application WO 98/18928), as these result in an enhanced immune response.

Microparticles are also preferred mucosal adjuvants. These are preferably derived from a poly(α-hydroxy acid), in particular, from a poly(lactide) ("PLA"), a copolymer of D,L-lactide and glycolide or glycolic acid, such as a poly(D,L-lactide-co-glycolide) ("PLG" or "PLGA"), or a copolymer of D,L-lactide and caprolactone. The microparticles may be derived from any of various polymeric starting materials which have a variety of molecular weights and, in the case of the copolymers such as PLG, a variety of lactide:glycolide ratios, the selection of which will be largely a matter of choice, depending in part on the coadministered antigen.

The SARS virus (inactivated or attenuated), viral antigens, antibodies or adjuvants of the invention may be entrapped within the microparticles, or may be adsorbed to them. Entrapment within PLG microparticles is preferred. PLG microparticles are discussed in further detail in Morris *et al.*, (1994), *Vaccine*, 12:5-11, in chapter 13 of *Mucosal Vaccines*, eds. Kiyono *et al.*, Academic Press 1996 (ISBN 012410587), and in chapters 16 & 18 of *Vaccine design: the subunit and adjuvant aproach*, eds. Powell & Newman, Plenum Press 1995 (ISBN 0-306-44867-X).

LT mutants may advantageously be used in combination with microparticle-entrapped antigen, resulting in significantly enhanced immune responses.

Aluminium compounds and MF59 are preferred adjuvants for parenteral use.

The composition may include an antibiotic.

The immunogenic compositions of the invention may be administered in a single dose, or as part of an administration regime. The regime may include priming and boosting doses, which may be administered mucosally, parenterally, or various combinations thereof.

The methods of the invention further comprise treating or preventing a SARS virus-related disease by administering to an animal a composition comprising an effective amount of the antibodies of the invention. An "effective amount" of the antibodies of the invention is an amount sufficient to provide passive immunization protection or treatment to the animal. Preferably, the antibodies of the invention are specific to the SARS viral antigen.

Methods of treatment may combine both immunogenic compositions and antibody compositions. Accordingly the invention comprises a method for treating or preventing a SARS virus-related disease comprising administering an immunogenic composition comprising an immunologically effective amount of a SARS viral antigen and administering an effective amount of antibodies specific to SARS viral antigen. The immunogenic composition and the antibodies may be administered together or separately. The invention further comprises a composition comprising an immunogenic composition comprising an immunologically effective amount of a SARS viral antigen and further comprising an effective amount of antibodies specific to a SARS viral antigen.

The SARS viral antigens and antibodies of the invention may also be administered in polynucleotide form. The SARS viral antigens and/or antibody proteins are then expressed *in vivo*.

The SARS viral antigens and the antibodies of the invention can also be delivered using one or more gene vectors, administered via nucleic acid immunization or the like using standard gene delivery protocols. Methods for gene delivery are known in the art. See, *e.g.*, US Patent Nos. 5,399,346, 5,580,859, 5,589,466. The constructs can be delivered (*e.g.*, injected) either subcutaneously, epidermally, intradermally, intramuscularly, intravenous, mucosally (such as nasally, rectally and vaginally), intraperitoneally, orally or combinations thereof. Intramuscular injection of 25μg plasmid DNA encoding spike antigens, in 200μl PBS pH 7.4, at weeks 0, 3 and 6, has been described for mice by Yang *et al.* (2004) *Nature* 428:561-564.

An exemplary replication-deficient gene delivery vehicle that may be used in the practice of the present invention is any of the alphavirus vectors, described in, for example, US Patent Nos. 6,342,372; 6,329,201 and International Publication WO 01/92552.

A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery systems. Selected sequences can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described (US Patent No. 5,219,740; Miller & Rosman, *BioTechniques* (1989) 7:980-990; Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa *et al.*, *Virology* (1991) 180:849-852; Burns *et al.*, *Proc. Natl. Acad. Sci. USA* (1993) 90:8033-8037; and Boris-Lawrie & Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109.

A number of adenovirus vectors have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol.* (1986) 57:267-274; Bett *et al.*, *J. Virol.* (1993) 67:5911-5921; Mittereder *et al.*, *Human Gene Therapy* (1994)

5:717-729; Seth *et al.*, *J. Virol.* (1994) 68:933-940; Barr *et al.*, *Gene Therapy* (1994) 1:51-58; Berkner, K.L. *BioTechniques* (1988) 6:616-629; and Rich *et al.*, *Human Gene Therapy* (1993) 4:461-476). Adenoviral delivery of codon-optimsed versions of the genes encoding SARS coronavirus structural antigens spike S1, membrane protein and nucleocapsid protein has been investigated in rhesus macaques and found to invoke a strong neutralizing antibody response (Gao *et al.* (2003) *Lancet* 362(9399):1895-1896).

Additionally, various adeno-associated virus (AAV) vector systems have been developed for gene delivery. AAV vectors can be readily constructed using techniques well known in the art. See, *e.g.*, US Patent Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 (published 23 January 1992) and WO 93/03769 (published 4 March 1993); Lebkowski *et al.*, *Molec. Cell. Biol.* (1988) 8:3988-3996; Vincent *et al.*, *Vaccines 90* (1990) (Cold Spring Harbor Laboratory Press); Carter, B.J. *Current Opinion in Biotechnology* (1992) 3:533-539; Muzyczka, N. *Current Topics in Microbiol. and Immunol.* (1992) 158:97-129; Kotin, R.M. *Human Gene Therapy* (1994) 5:793-801; Shelling and Smith, *Gene Therapy* (1994) 1:165-169; and Zhou *et al.*, *J. Exp. Med.* (1994) 179:1867-1875.

Another vector system useful for delivering polynucleotides, mucosally and otherwise, is the enterically administered recombinant poxvirus vaccines described by Small, Jr., P.A., *et al.* (US Patent No. 5,676,950, issued October 14, 1997, herein incorporated by reference) as well as the vaccinia virus and avian poxviruses. By way of example, vaccinia virus recombinants expressing the genes can be constructed as follows. The DNA encoding the SARS antigen or antibody or antibody coding sequence is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells that are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the coding sequences of interest into the viral genome. The resulting TK$^-$ recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver genes encoding the SARS viral antigens or antibodies of the invention. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an avipox vector is particularly desirable in human and other mammalian species since members of the avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia

viruses. See, *e.g.*, WO 91/12882; WO 89/03429; and WO 92/03545. Picornavirus-derived vectors can also be used. (See, *e.g.*, US Patent Nos. 5,614,413 and 6,063,384).

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael *et al.*, *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner *et al.*, *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery.

A vaccinia based infection/transfection system can be conveniently used to provide for inducible, transient expression of the coding sequences of interest (for example, a SARS viral antigen or antibody expression cassette) in a host cell. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA that is then translated into protein by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, *e.g.*, Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA* (1990) 87:6743-6747; Fuerst *et al.*, *Proc. Natl. Acad. Sci. USA* (1986) 83:8122-8126.

As an alternative approach to infection with vaccinia or avipox virus recombinants, or to the delivery of genes using other viral vectors, an amplification system can be used that will lead to high level expression following introduction into host cells. Specifically, a T7 RNA polymerase promoter preceding the coding region for T7 RNA polymerase can be engineered. Translation of RNA derived from this template will generate T7 RNA polymerase that in turn will transcribe more template. Concomitantly, there will be a cDNA whose expression is under the control of the T7 promoter. Thus, some of the T7 RNA polymerase generated from translation of the amplification template RNA will lead to transcription of the desired gene. Because some T7 RNA polymerase is required to initiate the amplification, T7 RNA polymerase can be introduced into cells along with the template(s) to prime the transcription reaction. The polymerase can be introduced as a protein or on a plasmid encoding the RNA polymerase. For a further discussion of T7 systems and their use for transforming cells, see, *e.g.*, International Publication No. WO 94/26911; Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130; Deng and Wolff, *Gene* (1994) 143:245-249; Gao *et al.*, *Biochem. Biophys. Res. Commun.* (1994) 200:1201-1206; Gao and Huang, *Nuc. Acids Res.* (1993) 21:2867-2872; Chen *et al.*, *Nuc. Acids Res.* (1994) 22:2114-2120; and US Patent No. 5,135,855.

The immunogenic compositions of the invention may further comprise diluents, such as water, saline, glycerol, ethanol, *etc.* Additionally, auxiliary substances, such as wetting or

emulsifying agents, pH buffering substances, and the like may be included in the immunogenic composition.

The immunogenic compositions used in the invention can be administered to an animal. Animals suitable for use in the methods of the invention include humans and other primates, including non-human primates such as chimpanzees, and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses, domestic animals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens, turkeys and other gallinaceous birds, ducks, geese and the like. Animals suitable for use in the invention can be of any age, including both adult and newborn. Transgenic animals can also be used in the invention.

The immunogenic compositions of the invention can be used to treat or prevent SARS virus-related diseases.

The compositions of the invention are preferably pharmaceutically acceptable and pharmacologically acceptable. In particularly, the compositions are preferably not biologically or otherwise undesirable, *i.e.*, the material may be administered to an individual in a formulation or composition without causing any undesirable biological effects or interacting in a deleterious manner with any of the components of the composition in which it is contained.

Pharmaceutically acceptable salts can also be used in compositions of the invention, for example, mineral salts such as hydrochlorides, hydrobromides, phosphates, or sulfates, as well as salts of organic acids such as acetates, proprionates, malonates, or benzoates. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those of skill in the art. Compositions of the invention can also contain liquids or excipients, such as water, saline, glycerol, dextrose, ethanol, or the like, singly or in combination, as well as substances such as wetting agents, emulsifying agents, or pH buffering agents. Liposomes can also be used as a carrier for a composition of the invention.

SARS specific reagents and analytical assays may be used in the manufacture and testing of the vaccines of the invention. Such analytical assays include, for example: 1) virus titration and plaque assays for quantitation of infectious virus particles, 2) a neutralization assay with constant virus and varying serum dilutions, 3) a two step RT-PCR system (Light Cycler-Roche) for detection of negative strand viral RNA, with the target sequence located within the N gene, providing highest possible sensitivity, and 4) ELISA and western blot assays for detection and qualification of viral proteins.

In addition, rabbit polyclonal antiserum has been generated to obtain antibody reagents (and demonstrate induction of neutralizing antibodies) against the SARS-CoV. A sample protocol for generating such reagents is set forth below. The virus is first cultivated in suitable

cell culture, such as Vero cells, and pelleted through a 20% sucrose (w/v) cushion. The pellet is then subjected to a glycerol potassium-tartrate gradient for further purification. The virus-containing fraction is then diluted and pelleted by ultracentrifugation. The pellet is then dissolved in PBS and the virus is inactivated with $C_3H_4O_2$ (beta-propiolactone, BPL). Two rabbits are immunized subcutaneously (SC) on day 0, 14, and 28 with $1x10^9$ inactivated viral particles mixed with IFA as adjuvant. Rabbits are bled on days 0 (pre-inoculation), 13, 28, and 35 (1 week after 3rd immunization). Sera obtained from this protocol were tested for their reactivity against SARS-CoV proteins in western blots and found to react with the major structural proteins spike (S), membrane (M), and nucleocapsid (N).

J. Emerging coronavirus vaccines

The SARS epidemic has lead to increased awareness of viral infections caused by coronaviruses. The vaccines of the invention may be adapted to prevent or treat emerging strains of coronavirus, including emerging strains of SARS virus.

The invention provides a vaccine comprising an inactivated (or killed) human coronavirus, an attenuated human coronavirus, a split human coronavirus preparation, or a recombinant or purified subunit formulation of one or more antigens from a human coronavirus, wherein the human coronavirus is not the SARS coronavirus. Optionally, the human coronavirus is not the 229E coronavirus. Optionally, the human coronavirus is not the OC43 coronavirus. Optionally, the human coronavirus is not the NL63 coronavirus. Thus the invention provides a vaccine as defined above, wherein the human coronavirus is not the SARS coronavirus, is not the 229E coronavirus, is not the OC43 coronavirus and is not the NL63 coronavirus. Such vaccines are useful for preventing and/or treating emerging human coronavirus infections.

The invention also provides a vaccine comprising: (a) an inactivated (or killed) human coronavirus, an attenuated human coronavirus, a split human coronavirus preparation, or a recombinant or purified subunit formulation of one or more antigens from a human coronavirus, wherein the human coronavirus is not the SARS coronavirus, as defined above; and (b) an inactivated (or killed) human coronavirus, an attenuated human coronavirus, a split human coronavirus preparation, or a recombinant or purified subunit formulation of one or more antigens from a human coronavirus, wherein the human coronavirus is the SARS coronavirus. Such vaccines are useful for preventing and/or treating both SARS and other human coronaviruses.

As well as providing vaccines comprising antigens from more than one type of coronavirus, the invention also provides vaccines comprising antigens from more than one strain of the same coronavirus *e.g.* different strains of the SARS coronavirus, or different strains of a coronavirus other than the SARS coronavirus. In one embodiment, the vaccine comprises antigens from at least two strains of coronavirus, or at least three strains of coronavirus. In one

embodiment, the vaccine comprises antigens from at least two types of coronavirus. In one embodiment, the vaccine comprises at least one antigen from each of the known types of coronaviruses (type I, type II and type III). Such vaccines follow the model of current influenza vaccines.

The selection of coronaviruses and/or coronavirus strains for use in vaccines of the invention can be based on various criteria. For instance, selection may be based on viruses and/or strains that have been detected in the geographical region (*e.g.* northern or southern hemisphere, a particular country, *etc.*) where the vaccine targeted. Selection may be based on the results of animal surveillance *e.g.* of viruses detected in cat populations. Selection may be based on the results of clinical surveillance *e.g.* of viruses detected in patients hospitalized with respiratory infection. Selection may be performed every year *e.g.* prior to winter. Vaccines may also be administered yearly, again following the model of current influenza vaccines.

Preferred vaccines are sufficiently immunogenic to provide a neutralizing immune response, and more preferably a protective and/or therapeutic immune response. Particularly preferred vaccines meet the efficacy requirements that may be specified by the WHO from time to time.

A preferred subunit antigen for inclusion in vaccines of the invention is a purified spike protein, more preferably in oligomeric (*e.g.* trimeric) form. The spike protein may or my not be cleaved *e.g.* into its S1 and S2 products.

The techniques disclosed above for selecting viruses and/or strains for production of vaccines can also be used to select appropriate viruses and/or strains from which HR1 and HR2 sequences can be obtained for providing therapeutic peptides, as disclosed above.

III. DIAGNOSTIC COMPOSITIONS AND METHODS OF THE INVENTION

The invention provides methods for detecting the SARS coronavirus. Detection in patient samples can be used to detect and diagnose infections by the virus. Detection in donated blood can be used to prevent inadvertent transmission of the virus during blood transplant procedures Detection methods fall into three main categories: detection of SARS virus nucleic acids; detection of SARS virus proteins; and detection of anti-SARS virus immune responses. The invention provides all such methods.

As used herein when referring to nucleotide sequences, particularly oligonucleotide probes and primers, "similar" sequences includes those sequences that are at least 90% identical to known SARSV genomic sequence and includes sequences that are at least 95 % identical, at least 99% identical and 100% identical to the SARSV genomic sequence over the length of the probe or primer.

As used herein, the term "target nucleic acid region" or "target nucleic acid" denotes a nucleic acid molecule with a "target sequence" to be amplified. The target nucleic acid may be either single-stranded or double-stranded and may include other sequences besides the target sequence, which may not be amplified. The term "target sequence" refers to the particular nucleotide sequence of the target nucleic acid which is to be amplified. The target sequence may include a probe-hybridizing region contained within the target molecule with which a probe will form a stable hybrid under desired conditions. The "target sequence" may also include the complexing sequences to which the oligonucleotide primers complex and be extended using the target sequence as a template. Where the target nucleic acid is originally single-stranded, the term "target sequence" also refers to the sequence complementary to the "target sequence" as present in the target nucleic acid. If the "target nucleic acid" is originally double-stranded, the term "target sequence" refers to both the plus (+) and minus (–) strands.

The term "primer" or "oligonucleotide primer" as used herein, refers to an oligonucleotide which acts to initiate synthesis of a complementary DNA strand when placed under conditions in which synthesis of a primer extension product is induced *i.e.* in the presence of nucleotides and a polymerization-inducing agent such as a DNA or RNA polymerase and at suitable temperature, pH, metal concentration, and salt concentration. The primer is preferably single-stranded for maximum efficiency in amplification, but may alternatively be double-stranded. If double-stranded, the primer is first treated to separate its strands before being used to prepare extension products. This denaturation step is typically effected by heat, but may alternatively be carried out using alkali, followed by neutralization. Thus, a "primer" is complementary to a template, and complexes by hydrogen bonding or hybridization with the template to give a primer/template complex for initiation of synthesis by a polymerase, which is extended by the addition of covalently bonded bases linked at its 3' end complementary to the template in the process of DNA synthesis.

As used herein, the term "probe" or "oligonucleotide probe" refers to a structure comprised of a polynucleotide, as defined above, that contains a nucleic acid sequence complementary to a nucleic acid sequence present in the target nucleic acid analyte. The polynucleotide regions of probes may be composed of DNA, and/or RNA, and/or synthetic nucleotide analogs. When an "oligonucleotide probe" is to be used in a 5' nuclease assay, such as the TaqMan™ technique, the probe will contain at least one fluorescer and at least one quencher which is digested by the 5' endonuclease activity of a polymerase used in the reaction in order to detect any amplified target oligonucleotide sequences. In this context, the oligonucleotide probe will have a sufficient number of phosphodiester linkages adjacent to its 5' end so that the 5' to 3' nuclease activity employed can efficiently degrade the bound probe to separate the fluorescers and quenchers.

When an oligonucleotide probe is used in the TMA technique, it will be suitably labeled, as described below.

It will be appreciated that the hybridizing sequences need not have perfect complementarity to provide stable hybrids. In many situations, stable hybrids will form where fewer than about 10% of the bases are mismatches, ignoring loops of four or more nucleotides. Accordingly, as used herein the term "complementary" refers to an oligonucleotide that forms a stable duplex with its "complement" under assay conditions, generally where there is about 90% or greater homology.

The terms "hybridize" and "hybridization" refer to the formation of complexes between nucleotide sequences which are sufficiently complementary to form complexes via Watson-Crick base pairing. Where a primer "hybridizes" with target (template), such complexes (or hybrids) are sufficiently stable to serve the priming function required by *e.g.* the DNA polymerase to initiate DNA synthesis.

Stringent hybridization conditions will typically include salt concentrations of less than about 1 M, more usually less than about 500 mM and preferably less than about 200 mM. Hybridization temperatures can be as low as 5°C, but are typically greater than 22°C, more typically greater than about 30°C, and preferably in excess of about 37°C. Longer fragments may require higher hybridization temperatures for specific hybridization. Other factors may affect the stringency of hybridization, including base composition and length of the complementary strands, presence of organic solvents and extent of base mismatching, and the combination of parameters used is more important than the absolute measure of any one alone. Other hybridization conditions which may be controlled include buffer type and concentration, solution pH, presence and concentration of blocking reagents to decrease background binding such as repeat sequences or blocking protein solutions, detergent type(s) and concentrations, molecules such as polymers which increase the relative concentration of the polynucleotides, metal ion(s) and their concentration(s), chelator(s) and their concentrations, and other conditions known in the art. Less stringent, and/or more physiological, hybridization conditions are used where a labeled polynucleotide amplification product cycles on and off a substrate linked to a complementary probe polynucleotide during a real-time assay which is monitored during PCR amplification such as a molecular beacon assay. Such less stringent hybridization conditions can also comprise solution conditions effective for other aspects of the method, for example reverse transcription or PCR.

As used herein, a "biological sample" refers to a sample of tissue, cells or fluid isolated from a subject, that commonly includes antibodies produced by the subject. Typical samples include but are not limited to, blood, plasma, serum, fecal matter, urine, bone marrow, bile, spinal fluid, lymph fluid, samples of the skin, secretions of the skin, respiratory, intestinal, and

genitourinary tracts, tears, saliva, sputum, mucous, milk, blood cells, organs, tissues, biopsies (*e.g.* lung, liver, kidney) and also samples of *in vitro* cell culture constituents including but not limited to conditioned media resulting from the growth of cells and tissues in culture medium *e.g.* recombinant cells, and cell components. Other samples that may be used for diagnosis include stool samples and nasopharyngeal aspirates.

The term "antibody" encompasses polyclonal and monoclonal antibody preparations, as well as preparations including hybrid antibodies, altered antibodies, chimeric antibodies and, humanized antibodies, as well as: hybrid (chimeric) antibody molecules (see, for example, Winter *et al.* (1991) *Nature* 349:293-299; and US Patent 4,816,567); $F(ab')_2$ and F(ab) fragments; Fv molecules (noncovalent heterodimers, see, for example, Inbar *et al.* (1972) *Proc Natl Acad Sci USA* 69:2659-2662; and Ehrlich *et al.* (1980) *Biochem* 19:4091-4096); single-chain Fv molecules (sFv) (*see, e.g.*, Huston *et al.* (1988) *Proc Natl Acad Sci USA* 85:5879-5883); oligobodies; dimeric and trimeric antibody fragment constructs; minibodies (see, *e.g.*, Pack *et al.* (1992) *Biochem* 31:1579-1584; Cumber *et al.* (1992) *J Immunology* 149B:120-126); humanized antibody molecules (see, *e.g.*, Riechmann *et al.* (1988) *Nature* 332:323-327; Verhoeyan *et al.* (1988) *Science* 239:1534-1536; and UK Patent Publication No. GB 2,276,169, published 21 September 1994); and, any functional fragments obtained from such molecules, wherein such fragments retain specific-binding properties of the parent antibody molecule.

As used herein, the term "monoclonal antibody" refers to an antibody composition having a homogeneous antibody population. The term is not limited regarding the species or source of the antibody, nor is it intended to be limited by the manner in which it is made. The term encompasses whole immunoglobulins.

Methods of making polyclonal and monoclonal antibodies are known in the art. Polyclonal antibodies are generated by immunizing a suitable animal, such as a mouse, rat, rabbit, sheep or goat, with an antigen of interest. In order to enhance immunogenicity, the antigen can be linked to a carrier prior to immunization. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Furthermore, the antigen may be conjugated to a bacterial toxoid, such as toxoid from diphtheria, tetanus, cholera, *etc.*, in order to enhance the immunogenicity thereof.

Rabbits, sheep and goats are preferred for the preparation of polyclonal sera when large volumes of sera are desired. These animals are good design choices also because of the availability of labeled anti-rabbit, anti-sheep and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the antigen in saline, preferably in an adjuvant such as

Freund's complete adjuvant ("FCA"), and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). The animal is generally boosted 2-6 weeks later with one or more injections of the antigen in saline, preferably using Freund's incomplete adjuvant ("FIA"). Antibodies may also be generated by in vitro immunization, using methods known in the art. Polyclonal antisera is then obtained from the immunized animal.

Monoclonal antibodies are generally prepared using the method of Kohler & Milstein (1975) *Nature* 256:495-497, or a modification thereof, as described above.

Nucleic acid detection methods

There are many well known methods of amplifying targeted sequences, such as the polymerase chain reaction (PCR), reverse transcription PCR (RT-PCR), the ligase chain reaction (LCR), the strand displacement amplification (SDA), and the nucleic acid sequence-based amplification (NASBA), transcription-mediated amplification (TMA) to name a few. These methods are described generally in the following references: (PCR) US Patents 4,683,195, 4,683,202, and 4,800,159; (RT-PCR) US patent 5,310,652, 5,322,770; (LCR) EP Application No., 320,308 published Jun. 14, 1989; (SDA) US Pat. Nos. 5,270,184, and 5,455,166 and "Empirical Aspects of Strand Displacement Amplification" by G. T. Walker in *PCR Methods and Applications*, 3(1):1-6 (1993), Cold Spring Harbor Laboratory Press; (TMA) US Patent No. 5,399,491, and (NASBA) "Nucleic Acid Sequence-Based Amplification (NASBA™)" by L. Malek *et al.*, Ch. 36 in *Methods in Molecular Biology*, Vol. 28: Protocols for Nucleic Acid Analysis by Nonradioactive Probes, 1994 Ed. P. G. Isaac, Humana Press, Inc., Totowa, N.J. PCR methods may include variations that permit quantitation of the target sequence, for example, by real time PCR analysis (e.g., as described in US patents 5,210,015, 5,487,972, 5,994,056, 6,171,785 inter alia). (Each of the above references are hereby incorporated by reference).

One embodiment of the method of the invention for detecting the presence of SARS virus in a sample comprises providing a sample suspected of containing a SARS virus nucleic acid target, amplifying a template sequence contained within said SARS virus nucleic acid target by any known technique of nucleic acid amplification, including any of those mentioned herein, using the oligonucleotide primers described herein, particularly those primers comprising the kits described herein, and detecting the amplified template sequence, wherein the presence of the amplified template sequence indicates the presence of SARS virus in said sample.

Amplification techniques generally involve the use of two primers. Where a target sequence is single-stranded, the techniques generally involve a preliminary step in which a complementary strand is made in order to give a double-stranded target. The two primers hybridize to different strands of the double-stranded target and are then extended. The extended products can serve as targets for further rounds of hybridization/extension. The net effect is to amplify a template sequence within the target, the 5' and 3' termini of the template being defined

by the locations of the two primers in the target. As an alternative, if one or both of the primers contains a promoter sequence then the target can be amplified (by transcription) using a RNA polymerase (as in TMA).

The present invention provides methods and kits for amplifying and/or detecting a template or target sequence in the SARSV viral nucleic acid. The invention provides a kit comprising primers for amplifying a template sequence contained within a SARSV nucleic acid target, the kit comprising a first primer and a second primer, wherein the first primer comprises a sequence substantially complementary to a portion of said template sequence and the second primer comprises a sequence substantially complementary to a portion of the complement of said template sequence, wherein the sequences within said primers which have substantial complementarity define the termini of the template sequence to be amplified.

Kits of the invention may further comprise a probe which is substantially complementary to the template sequence and/or to its complement and which can hybridize thereto. This probe can be used in a hybridization technique to detect amplified template, or to isolate (*i.e.* "capture) the amplified template or the original target nucleic acid.

Kits of the invention may further comprise primers and/or probes for generating and detecting an internal standard, in order to aid quantitative measurements (*e.g* Fille *et al.* 1997 *Biotechniques* 23:34-36).

Kits of the invention may further comprise a DNA polymerase, which will generally be a thermostable DNA polymerase where a non-isothermal amplification process is to be used. The kits may also comprise supplies of dNTPs, a magnesium salt (*e.g.* $MgCl_2$), buffer solutions, *etc.*

Kits of the invention may comprise more than one pair of primers (*e.g.* for nested amplification), and one primer may be common to more than one primer pair. The kit may also comprise more than one probe.

Oligomer Probes and Primers

In connection with the nucleic acid detection methods of the present invention described above, oligomers having sequence similarity, or complementarity, to the SARSV genome are useful. The SARSV genome sequences mentioned herein may be used to produce probes and primers which can be used in assays for the detection of nucleic acids in test samples. The probes may be designed from conserved nucleotide regions of the polynucleotides of interest or from non-conserved nucleotide regions of the polynucleotide of interest. The design of such probes for optimization in assays is within the skill of those of ordinary skill in the art. Generally, nucleic acid probes are developed from non-conserved or unique regions when maximum specificity is desired, and nucleic acid probes are developed from conserved regions when assaying for nucleotide regions that are closely related to, for example, different members of a multi-gene family or in related species like mouse and man.

Using as a basis the SARSV genome which can be found as described herein, and/or preferably conserved regions of the SARSV genome, and/or the particularly described primer and probe sequences as disclosed herein, oligomers of approximately 8 nucleotides or more can be prepared which hybridize with the positive strand(s) of SARSV RNA or its complement, as well as to SARSV cDNAs. These oligomers can serve as probes for the detection (including isolation and/or labeling) of polynucleotides which contain SARSV nucleotide sequences, and/or as primers for the transcription and/or replication of targeted SARSV sequences. The oligomers contain a targeting polynucleotide sequence, which is comprised of nucleotides which are complementary to a target SARSV nucleotide sequence; the sequence is of sufficient length and complementarity with the SARSV sequence to form a duplex which has sufficient stability for the purpose intended. For example, if the purpose is the isolation, via immobilization, of an analyte containing a target SARSV sequence, the oligomers would contain a polynucleotide region which is of sufficient length and complementarity to the targeted SARSV sequence to afford sufficient duplex stability to immobilize the analyte on a solid surface, via its binding to the oligomers, under the isolation conditions. For example, also, if the oligomers are to serve as primers for the transcription and/or replication of target SARSV sequences in an analyte polynucleotide, the oligomers would contain a polynucleotide region of sufficient length and complementarity to the targeted SARSV sequence to allow the polymerizing agent to continue replication from the primers which are in stable duplex form with the target sequence, under the polymerizing conditions. For example, also, if the oligomers are to be used as label probes, or are to bind to multimers, the targeting polynucleotide region would be of sufficient length and complementarity to form stable hybrid duplex structures with the label probes and/or multimers to allow detection of the duplex. The oligomers may contain a minimum of about 4 contiguous nucleotides which are complementary to targeted SARSV sequence; usually the oligomers will contain a minimum of about 8 contiguous nucleotides which are complementary to the targeted SARSV sequence, and preferably will contain a minimum of about 14, 15, 16, 17, 18, 19, 20,21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 contiguous nucleotides and up to about 50, 75, 100, 200 contiguous nucleotides or more, which are complementary to the targeted SARSV sequence.

Typically, for use in the amplification based methods (for example, PCR, RT-PCR, TMA) oligomers will be used as primer sets such that one member of the primer set has sequence similarity or complementarity to a more conserved (among coronaviruses) portion of the SARSV genome and the other member of the primer set has sequence similarity or complementarity to a less conserved portion. The primer sets can be used to amplify the target region in ways that are well known in the art. Typically, the 5' untranslated region (5'UTR) and the 3' untranslated region (3'UTR) are among the most conserved regions. Figure 8 shows an alignment of the 5'UTR of several coronaviruses. Figure 10 shows an alignment of the 3'UTR of several

coronaviruses. Figures 9 and 11 show the sequences of preferred primers for amplification of the 5'UTR and 3'UTR, respectively. Other primers and probes can readily be designed based on the sequence alignments provided herein.

The oligomer, however, need not consist only of the sequence which is complementary to the targeted SARSV sequence. It may contain in addition, nucleotide sequences (*e.g.*promoters) or other moieties which are suitable for the purposes for which the oligomers are used. For example, if the oligomers are used as primers for the amplification of SARSV sequences via, for example, PCR, they may contain sequences which, when in duplex, form restriction enzyme sites which facilitate the cloning of the amplified sequences. For example, also, if the oligomers are to be used as "capture probes" in hybridization assays, they would contain in addition a binding partner which is coupled to the oligomer containing the nucleotide sequence which is complementary to the targeted SARSV sequence. Other types of moieities or sequences which are useful of which the oligomers may be comprised or coupled to, are those which are known in the art to be suitable for a variety of purposes, including the labeling of nucleotide probes.

Table 4 (SEQ ID NOS: 1021-6020) shows forward and reverse primers that are useful for nucleic acid amplification of SARSV for diagnostic and screening methods.

Preferred primers and probes for SARS nucleic acid detection for diagnostic and screening are SEQ ID NOS: 7332-7336 (forward primers), SEQ ID NOS: 7337-7341 (reverse primers) and SEQ ID NOS: 7342-7352 (probes). These primers and probes are useful for detection of sequences in the 3' UTR.

Any of the above forward primers may be used in combination with any of the above reverse primers for amplification of SARSV nucleic acid. The amplified product may be detected (or captured) with any of the above probes. Particularly preferred combinations of forward and reverse primers and the probes for detecting the amplified product include: Forward SEQ ID NO: 7332 with reverse SEQ ID NO: 7337, 7338, 7339 or 7341 and probe SEQ ID NO: 7342; forward SEQ ID NO: 7333 or 7334 with reverse SEQ ID NO: 7340 and any of probes SEQ ID NO: 7343-7351; Forward SEQ ID NO: 7335 and reverse SEQ ID NO: 7340 or 7341 and any of probes SEQ ID NO: 7342-7352. Other combinations of forward and reverse primers and appropriate probes can readily be determined by those skilled in the art from the above information.

Additional preferred primers and probes for SARS nucleic acid detection for diagnostic and screening are SEQ ID NOS: 7353-7362 (forward primers), SEQ ID NOS: 7363-7373 (reverse primers) and SEQ ID NOS: 7374-7385 (probes). The primers and probes are useful for detection of sequences in the 5' UTR.

The above primers may be used in combination for amplification of SARSV nucleic acid as follows: any of forward primers SEQ ID NO: 7353-7356 with any of reverse primers SEQ ID

NO: 7363-7366, 7368 and the amplified product detected (or captured) with probes SEQ ID NO: 7374; any of forward primers SEQ ID NO: 7357-7362 with any of reverse primers SEQ ID NO: 7367, 7369-7373 and the amplified products detected (or captured) with any of probes SEQ ID NO: 7375-7385. Particularly preferred combinations of forward and reverse primers and probes are: Forward primers SEQ ID NO: 7353-7356 with any of reverse primers SEQ ID NO: 7363-7366 and probes SEQ ID NO: 7374; forward primers SEQ ID NO: 7357-7358 with reverse primers SEQ ID NO: 7367, 7369 and probes SEQ ID NO: 7375 or 7376; Forward primers SEQ ID NO: 7357-7359 with reverse primers SEQ ID NO: 7367, 7369 or 7370 and probe SEQ ID NO: 7375 or 7376. More preferred are combinations of SEQ ID NO: 7353 or 7354 with SEQ ID NO: 7363 or 7364 and probe SEQ ID NO: 7374. Other combinations of forward and reverse primers and appropriate probes can readily be determined by those skilled in the art from the above information. A particularly conserved octanucleotide sequence (SEQ ID NO: 7386) occurs in the 3'UTR of SARS (approximately 70-80 bases from the 3' end) and of several other Coronaviruses that may be particularly useful in identifying SARSV. Primers including in this region are preferably combined with reverse primers from regions of sequence that are more specific for SARS.

In addition to the above, the intergenic sequence (IS) that is characteristic of Coronavirus has been identified in SARSV (see above). The IS minimally comprises the sequence ACGAAC (SEQ ID NO: 7293) which occurs upstream of each open reading frame (ORF) in the viral genome. The 5'UTR which includes the IS is spliced onto the 5' end of each viral mRNA at or adjacent to the site of the IS. Thus, primers comprising the IS or its complement are useful for amplifying viral nucleic acids, including cDNA made from the viral RNAs. The invention thus comprises a set of primers in which one primer comprises ACGAAC (SEQ ID NO: 7293) or its complement (SEQ ID NO: 7387) and one primer comprises any appropriate sequence from the SARS genome, or a complementary sequence. Useful probes for detecting and/or capturing the viral RNAs or cDNA made from the viral RNAs may also comprise the IS sequence, or its complement, described above.

One set of primers for amplification of SARS sequences, particularly by RT-PCR, uses SEQ ID NOs 6562, 6563, 6564 and 6565. Of these, 6562 & 6564 are sense primers and 6563 & 6565 are antisense primers. Primers SEQ ID NOS: 6562 & 6565 may be used in a first amplification, with a second nested amplification being performed using primers SEQ ID NOS: 6563 & 6564. In some embodiments of the invention, these four primers are excluded.

One kit for amplification and detection of SARS sequences, particularly by RT-PCR, uses SEQ ID NOs 6567 & 6568 as primers, and SEQ ID NO 6566 as a probe (typically labeled *e.g.* with TAMRA and/or FAM) for the amplified sequence. In some embodiments of the invention, these primers and probe are excluded.

One kit for amplification and detection of SARS sequences, particularly by RT-PCR, uses SEQ ID NOs 7395 & 6568 as primers, and SEQ ID NO 6566 as a probe (typically labeled *e.g.* with TAMRA and/or FAM) for the amplified sequence. In some embodiments of the invention, these primers and probe are excluded.

One kit for amplification of SARS sequences, particularly the nucleocapsid gene, uses SEQ ID NOs 6560 & 6561 as primers. In some embodiments of the invention, these primers are excluded.

One kit for amplification of SARS sequences uses SEQ ID NOs 6496, 6497, 6562, 6563, 6564 & 6565 as primers. In some embodiments of the invention, these primers are excluded.

One kit for amplification of SARS sequences uses SEQ ID NOs 6562, 6563, 6564 & 6565 as primers. In some embodiments of the invention, these primers are excluded.

One kit for amplification of SARS sequences uses SEQ ID NOs 6500, 6501, 6502 & 6503 as primers. In some embodiments of the invention, these primers are excluded.

One kit for amplification of SARS sequences uses SEQ ID NOs 6496, 6497, 6500, 6501, 6502, 6503, 6562, 6563, 6564 & 6565 as primers. In some embodiments of the invention, these primers are excluded.

One kit for amplification and detection of SARS sequences, particularly by realtime (*e.g.* TaqMan™) PCR, uses SEQ ID NOs 6567 & 6568 as primers, and SEQ ID NO 6566 as a probe (typically labeled *e.g.* with TAMRA and/or FAM) for the amplified sequence. In some embodiments of the invention, these primers and probe are excluded.

One kit for amplification and detection of SARS sequences, particularly by realtime (*e.g.* TaqMan™) PCR, uses SEQ ID NOs 7395 & 6568 as primers, and SEQ ID NO 6566 as a probe (typically labeled *e.g.* with TAMRA and/or FAM) for the amplified sequence. In some embodiments of the invention, these primers and probe are excluded.

One kit for amplification and detection of SARS sequences uses SEQ ID NOs 6562, 6565 and 6568 as primers, and SEQ ID NOs 7396 and 7397 as probes (typically labeled *e.g.* with TAMRA and/or FAM) for the amplified sequence. In some embodiments of the invention, these primers and probe are excluded.

One kit for amplification and detection of SARS sequences uses an oligonucleotide comprising SEQ ID NO: 9780 as a forward primer, an oligonucleotide comprising SEQ ID NO: 9781 as a reverse primer, and an oligonucleotide comprising SEQ ID NO: 9782 as a probe.

Preferred sequences for use with RT-PCR and LightCycler analysis include SEQ ID NOs 6562, 6568, 6565, 7396 & 7397. In some embodiments of the invention, these primers and probe are excluded.

The preparation of the oligomers is by means known in the art, including, for example, by methods which include excision, transcription, or chemical synthesis. The target sequences

and/or regions of the genome which are selected to which the targeting polynucleotides of the oligomers are complementary depend upon the purpose. For example, if the goal is to screen for the presence of SARSV in biological samples (*e.g.* blood, respiratory material, liver, lung), the preferred oligomers would be used as probes and/or primers, and would hybridize to conserved regions of the SARSV genome. Some of the conserved regions of the SARSV genome to which the oligomers may bind are described herein, for example, 5'UTR and 3'UTR.

In the basic nucleic acid hybridization assay, single-stranded analyte nucleic acid (either DNA or RNA) is hybridized to a nucleic acid probe, and resulting duplexes are detected. The probes for SARSV polynucleotides (natural or derived) are a length which allows the detection of unique viral sequences by hybridization. While 6-8 nucleotides may be a workable length, sequences of 10-12 nucleotides are preferred, and about 13, 14, 15, 16, 17, 18, 19, 20, or 21 or more nucleotides or more appears optimal. Preferably, these sequences will derive from regions which lack heterogeneity. These probes can be prepared using routine methods, including automated oligonucleotide synthetic methods. Among useful probes, for example, are those derived from less conserved regions of the SARSV genome. Regions of the genome that are typically less conserved can be readily ascertained from the sequence alignments provided herein, as well as by any other well known techniques. A complement to any unique portion of the SARSV genome will be satisfactory. For use as probes, complete complementarity is desirable, though it may be unnecessary as the length of the fragment is increased.

For use of such probes as agents to detect the presence of SARSV polynucleotides (for example in screening for contaminated blood or for diagnosing infected individuals), the biological sample to be analyzed, such as, without limitation, blood, serum, lung, liver, mucous, kidney, saliva, or sputum, may be treated, if desired, to extract the nucleic acids contained therein. The resulting nucleic acid from the sample may be subjected to gel electrophoresis or other size separation techniques; alternatively, the nucleic acid sample may be dot blotted without size separation. In order to form hybrid duplexes with the targeting sequence of the probe, the targeted region of the analyte nucleic acid must be in single stranded form. Where the sequence is naturally present in single stranded form, denaturation will not be required. However, where the sequence is present in double stranded form, the sequence will be denatured. Denaturation can be carried out by various techniques known in the art. Subsequent to denaturation, the analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the analyte, and the resulting duplexes containing the probe(s) are detected.

Detection of the resulting duplex, if any, is usually accomplished by the use of labeled probes; alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labeled, either directly or indirectly. Suitable labels, and methods for labeling

probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (*e.g.*, nick translation or kinasing), biotin, fluorescent groups, chemiluminescent groups (*e.g.*, dioxetanes, particularly triggered dioxetanes), enzymes, antibodies, and the like.

The region of the probes which are used to bind to the analyte can be made completely complementary to the SARSV genome. Therefore, usually high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency should only be used if the probes are complementary to regions of the viral genome which lack heterogeneity. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, length of time, and concentration of formamide. These factors are outlined in, for example, Maniatis T. (1982).

Variations of this basic scheme which are known in the art, including those which facilitate separation of the duplexes to be detected from extraneous materials and/or which amplify the signal from the labeled moiety, may also be used. A number of these variations are reviewed in, for example: Matthews & Kricka (1988), *Analytical Biochemistry* 169:1; Landegren *et al.* (1988), Science 242:229; and Mittlin (1989), *Clinical Chem.* 35:1819. These and the following publications describing assay formats are hereby incorporated by reference herein. Probes suitable for detecting SARSV in these assays are comprised of sequences which hybridize with target SARSV polynucleotide sequences to form duplexes with the analyte strand, wherein the duplexes are of sufficient stability for detection in the specified assay system.

A suitable variation is, for example, one which is described in US Pat. No. 4,868,105, issued Sep. 9, 1989, and in EPO Publication No. 225,807 (published Jun. 16, 1987). These publications describe a solution phase nucleic acid hybridization assay in which the analyte nucleic acid is hybridized to a labeling probe set and to a capturing probe set. The probe-analyte complex is coupled by hybridization with a solid-supported capture probe that is complementary to the capture probe set. This permits the analyte nucleic acid to be removed from solution as a solid phase complex. Having the analyte in the form of a solid phase complex facilitates subsequent separation steps in the assay. The labeling probe set is complementary to a labeled probe that is bound through hybridization to the solid phase/analyte complex.

The polymerase chain reaction (PCR) is a technique for amplifying a desired nucleic acid sequence (target) contained in a nucleic acid or mixture thereof. In PCR, a pair of primers are employed in excess to hybridize to the complementary strands of the target nucleic acid. The primers are each extended by a polymerase using the target nucleic acid as a template. The extension products become target sequences themselves, following dissociation from the original target strand. New primers then are hybridized and extended by a polymerase, and the cycle is

repeated to geometrically increase the number of target sequence molecules. PCR is disclosed in US Pat. Nos. 4,683,195 and 4,683,202, which are incorporated herein by reference.

The Ligase Chain Reaction (LCR) is an alternate method for nucleic acid amplification. In LCR, probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand, and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3' hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. This technique is described more completely in EP-A-320 308 to K. Backman published Jun. 16, 1989 and EP-A-0439182 to K. Backman *et al.*, published Jul. 31, 1991, both of which are incorporated herein by reference.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in US Pat. No. 5,322,770, which is incorporated herein by reference; or reverse transcribe mRNA into cDNA followed by asymmetric gap ligase chain reaction (RT-AGLCR) as described by R. L. Marshall *et al.*, PCR Methods and Applications 4:80-84 (1994), which also is incorporated herein by reference.

TMA is described in detail in, *e.g.*, US Patent No. 5,399,491, the disclosure of which is incorporated herein by reference in its entirety. In one example of a typical assay, an isolated nucleic acid sample, suspected of containing a SARSV target sequence, is mixed with a buffer concentrate containing the buffer, salts, magnesium, nucleotide triphosphates, primers, dithiothreitol, and spermidine. The reaction is optionally incubated at about 100°C for approximately two minutes to denature any secondary structure. After cooling to room temperature, reverse transcriptase, RNA polymerase, and RNAse H are added and the mixture is incubated for two to four hours at 37°C. The reaction can then be assayed by denaturing the product, adding a probe solution, incubating 20 minutes at 60°C, adding a solution to selectively hydrolyze the unhybridized probe, incubating the reaction six minutes at 60°C, and measuring the remaining chemiluminescence in a luminometer.

Generally, TMA includes the following steps: (a) isolating nucleic acid, including RNA, from the biological sample of interest suspected of being infected with SARSV; and (b) combining into a reaction mixture (i) the isolated nucleic acid, (ii) first and second oligonucleotide primers, the first primer having a complexing sequence sufficiently complementary to the 3' terminal portion of an RNA target sequence, if present (for example the (+) strand), to complex therewith, and the second primer having a complexing sequence sufficiently complementary to the 3' terminal portion of the target sequence of its complement (for example, the (-) strand) to complex therewith, wherein the first oligonucleotide further comprises a sequence 5' to the complexing sequence which includes a promoter, (iii) a reverse transcriptase or RNA and DNA dependent DNA polymerases, (iv) an enzyme activity which selectively degrades the RNA strand of an RNA-DNA complex (such as an RNAse H) and (v) an RNA polymerase which recognizes the promoter.

The components of the reaction mixture may be combined stepwise or at once. The reaction mixture is incubated under conditions whereby an oligonucleotide/target sequence is formed, including DNA priming and nucleic acid synthesizing conditions (including ribonucleotide triphosphates and deoxyribonucleotide triphosphates) for a period of time sufficient to provide multiple copies of the target sequence. The reaction advantageously takes place under conditions suitable for maintaining the stability of reaction components such as the component enzymes and without requiring modification or manipulation of reaction conditions during the course of the amplification reaction. Accordingly, the reaction may take place under conditions that are substantially isothermal and include substantially constant ionic strength and pH. The reaction conveniently does not require a denaturation step to separate the RNA-DNA complex produced by the first DNA extension reaction.

Suitable DNA polymerases include reverse transcriptases, such as avian myeloblastosis virus (AMV) reverse transcriptase (available from, *e.g.*, Seikagaku America, Inc.) and Moloney murine leukemia virus (MMLV) reverse transcriptase (available from, *e.g.*, Bethesda Research Laboratories).

Promoters or promoter sequences suitable for incorporation in the primers are nucleic acid sequences (either naturally occurring, produced synthetically or a product of a restriction digest) that are specifically recognized by an RNA polymerase that recognizes and binds to that sequence and initiates the process of transcription whereby RNA transcripts are produced. The sequence may optionally include nucleotide bases extending beyond the actual recognition site for the RNA polymerase which may impart added stability or susceptibility to degradation processes or increased transcription efficiency. Examples of useful promoters include those which are recognized by certain bacteriophage polymerases such as those from bacteriophage

T3, T7 or SP6, or a promoter from *E. coli*. These RNA polymerases are readily available from commercial sources, such as New England Biolabs and Epicentre.

Some of the reverse transcriptases suitable for use in the methods herein have an RNAse H activity, such as AMV reverse transcriptase. It may, however, be preferable to add exogenous RNAse H, such as *E. coli* RNAse H, even when AMV reverse transcriptase is used. RNAse H is readily available from, *e.g.*, Bethesda Research Laboratories.

The RNA transcripts produced by these methods may serve as templates to produce additional copies of the target sequence through the above-described mechanisms. The system is autocatalytic and amplification occurs autocatalytically without the need for repeatedly modifying or changing reaction conditions such as temperature, pH, ionic strength or the like.

Detection may be done using a wide variety of methods, including direct sequencing, hybridization with sequence-specific oligomers, gel electrophoresis and mass spectrometry. these methods can use heterogeneous or homogeneous formats, isotopic or nonisotopic labels, as well as no labels at all.

Suitable labeling moieties for attachment to primers and/or to probes used in methods of the invention include, but are not limited to: 5-FAM (also called 5-carboxyfluorescein; also called Spiro(isobenzofuran-1(3H), 9'-(9H)xanthene)-5-carboxylic acid,3',6'-dihydroxy-3-oxo-6-carboxyfluorescein); 5-Hexachloro-Fluorescein ([4,7,2',4',5',7'-hexachloro-(3',6'-dipivaloylfluoresceinyl)-6-carboxylic acid]); 6-Hexachloro-Fluorescein ([4,7,2',4',5',7'-hexachloro-(3',6'-dipivaloylfluoresceinyl)-5-carboxylic acid]); 5-Tetrachloro-Fluorescein ([4,7,2',7'-tetrachloro-(3',6'-dipivaloylfluoresceinyl)-5- carboxylic acid]); 6-Tetrachloro-Fluorescein ([4,7,2',7'-tetrachloro-(3',6'-dipivaloylfluoresceinyl)-6- carboxylic acid]); tetramethylrhodamines (TAMRA), including (i) 5-TAMRA (5-carboxytetramethylrhodamine; Xanthylium, 9-(2,4-dicarboxyphenyl)-3,6- bis(dimethylamino) and (ii) 6-TAMRA (6-carboxytetramethylrhodamine; Xanthylium, 9-(2,5-dicarboxyphenyl)-3,6- bis(dimethylamino); EDANS (5-((2-aminoethyl)amino)naphthalene- 1-sulfonic acid) ; 1,5-IAEDANS (5-((((2-iodoacetyl)amino)ethyl) amino)naphthalene-1-sulfonic acid); DABCYL (4-((4-(dimethylamino)phenyl) azo)benzoic acid); Cy5 (Indodicarbocyanine-5); Cy3 (Indodicarbocyanine-3); and BODIPY™ FL (4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-propionic acid). Labelling of probes with both FAM (*e.g.* at 5') and TAMRA (*e.g.* at 3') is preferred.

Nucleic acids of the invention may be used in solution or may be bound to a solid matrix or support *e.g.* in the format of a DNA array,

As is readily apparent, design of the assays described herein are subject to a great deal of variation, and many formats are known in the art. The above descriptions are merely provided as

guidance and one of skill in the art can readily modify the described protocols, using techniques well known in the art.

One 302nt amplicon of the SARS virus is known as "BNI-1" (SEQ ID NO: 9927). It was sequenced at the Bernhard Nocht Institute, Hamburg, Germany. In April 2003 the BNI-1 sequence was published on the WHO website (*http://www.who.int/csr/sars/primers/en/*) and in Dorsten *et al.*, "Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome", *New England Journal of Medicine*, published online at *http://www.nejm.org*. Both references are incorporated herein by reference in their entirety. Some embodiments of the invention do not encompass a nucleic acid consisting of SEQ ID NO: 9927. Some other embodiments of the invention do not encompass a nucleic acid comprising SEQ ID NO: 9927. Some embodiments of the invention do not encompass a polypeptide consisting of any one of SEQ ID NO^S: 9928 to 9959. Some other embodiments of the invention do not encompass a nucleic acid comprising any one of SEQ ID NO^S: 9928 to 9959. Some embodiments of the invention are not subject to these exclusions.

Immunoassays

The present invention utilizes various immunoassay techniques for identifying individuals exposed to SARSV and/or biological samples containing SARSV antigens or antibodies to SARSV.

Immunoassay Formats

The SARSV antigens may be employed in virtually any assay format that employs a known antigen to detect antibodies. A common feature of all of these assays is that the antigen is contacted with biological sample suspected of containing SARSV antibodies under conditions that permit the antigen to bind to any such antibody present in the component. Such conditions will typically be physiologic temperature, pH and ionic strength using an excess of antigen. The incubation of the antigen with the specimen is followed by detection of immune complexes comprised of the antigen. Alternatively, anti-SARSV antibodies may be employed to detect the presence of SARSV antigens in a biological sample. Combination antigen/antibody assays are also contemplated; for example, as described for HCV detection in US patent 6,630,298.

Design of the immunoassays is subject to a great deal of variation, and many formats are known in the art. Protocols may, for example, use solid supports, or immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, enzymatic, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the immune complex are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

The immunoassay may be, without limitation, in a heterogeneous or in a homogeneous format, and of a standard or competitive type. In a heterogeneous format, the polypeptide is typically bound to a solid matrix or support to facilitate separation of the sample from the polypeptide after incubation. Examples of solid supports that can be used are nitrocellulose (*e.g.*, in membrane or microtiter well form), polyvinyl chloride (*e.g.*, in sheets or microtiter wells), polystyrene latex (*e.g.*, in beads or microtiter plates, polyvinylidine fluoride, diazotized paper, nylon membranes, microchips, high or low density biochips, recombinant immunoassays (RIBA), microfluidity devices, micromagnetic beads, activated beads, and Protein A beads. For example, Dynatech Immunlon or Immunlon 2 microtiter plates or 0.25 inch polystyrene beads (Precision Plastic Ball) can be used in the heterogeneous format. The solid support containing the antigenic polypeptides is typically washed after separating it from the test sample, and prior to detection of bound antibodies. Both standard and competitive formats are known in the art.

In a homogenous format, the test sample is incubated with the combination of antigens in solution. For example, it may be under conditions that will precipitate any antigen-antibody complexes which are formed. Both standard and competitive formats for these assays are known in the art.

In a standard format, the amount of SARSV antibodies in the antibody-antigen complexes is directly monitored. This may be accomplished by determining whether labeled anti-xenogeneic (*e.g.*, anti-human) antibodies which recognize an epitope on anti-SARSV antibodies will bind due to complex formation. In a competitive format, the amount of SARSV antibodies in the sample is deduced by monitoring the competitive effect on the binding of a known amount of labeled antibody (or other competing ligand) in the complex.

Complexes formed comprising anti-SARSV antibody (or in the case of competitive assays, the amount of competing antibody) are detected by any of a number of known techniques, depending on the format. For example, unlabeled SARSV antibodies in the complex may be detected using a conjugate of antixenogeneic Ig complexed with a label, (*e.g.*, an enzyme label).

In an immunoprecipitation or agglutination assay format the reaction between the SARSV antigens and the antibody forms a network that precipitates from the solution or suspension and forms a visible layer or film of precipitate. If no anti-SARSV antibody is present in the test specimen, no visible precipitate is formed.

There are at least three specific types of particle agglutination (PA) assays. These assays are used for the detection of antibodies to various antigens when coated to a support. One type of this assay is the hemagglutination assay using red blood cells (RBCs) that are sensitized by passively adsorbing antigen (or antibody) to the RBC. The addition of specific antigen antibodies present in the body component, if any, causes the RBCs coated with the purified antigen to agglutinate.

To eliminate potential non-specific reactions in the hemagglutination assay, two artificial carriers may be used instead of RBC in the PA. The most common of these are latex particles. However, gelatin particles may also be used. The assays utilizing either of these carriers are based on passive agglutination of the particles coated with purified antigens.

The SARSV antigens will typically be packaged in the form of a kit for use in these immunoassays. The kit will normally contain in separate containers the native SARSV antigen, control antibody formulations (positive and/or negative), labeled antibody when the assay format requires same and signal generating reagents (*e.g.*, enzyme substrate) if the label does not generate a signal directly. The native SARSV antigen may be already bound to a solid matrix or separate with reagents for binding it to the matrix. Instructions (*e.g.*, written, tape, CD-ROM, *etc.*) for carrying out the assay usually will be included in the kit.

Immunoassays that utilize the native SARSV antigen are additionally useful in screening blood for the preparation of a supply from which potentially infective SARSV is lacking. The method for the preparation of the blood supply comprises the following steps. Reacting a body component, preferably blood or a blood component, from the individual donating blood with native SARSV antigen to allow an immunological reaction between SARSV antibodies, if any, and the SARSV antigen. Detecting whether anti-SARSV antibody--SARSV antigen complexes are formed as a result of the reacting. Blood contributed to the blood supply is from donors that do not exhibit antibodies to the native SARSV antigens.

Production of Antibodies

As explained above, the assay may utilize various antibodies which may be bound to a solid support, and that detect antigen or antigen/antibody complexes formed when SARSV infection is present in the sample. These antibodies may be polyclonal or monoclonal antibody preparations, monospecific antisera, human antibodies, or may be hybrid or chimeric antibodies, such as humanized antibodies, altered antibodies, $F(ab')_2$ fragments, F(ab) fragments, Fv fragments, single-domain antibodies, dimeric or trimeric antibody fragment constructs, minibodies, or functional fragments thereof which bind to the antigen in question.

Antibodies are produced using techniques well known to those of skill in the art and disclosed in, for example, US Pat. Nos. 4,011,308; 4,722,890; 4,016,043; 3,876,504; 3,770,380; and 4,372,745. For example, polyclonal antibodies are generated by immunizing a suitable animal, such as a mouse, rat, rabbit, sheep or goat, with an antigen of interest. In order to enhance immunogenicity, the antigen can be linked to a carrier prior to immunization. Such carriers are well known to those of ordinary skill in the art. Immunization is generally performed by mixing or emulsifying the antigen in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). The animal is generally boosted 2-6 weeks later with one or more injections

of the antigen in saline, preferably using Freund's incomplete adjuvant. Antibodies may also be generated by *in vitro* immunization, using methods known in the art. Polyclonal antiserum is then obtained from the immunized animal.

Monoclonal antibodies are generally prepared using the method of Kohler & Milstein (1975) *Nature* 256:495-497, or a modification thereof, as described above.

As explained above, antibody fragments which retain the ability to recognize the antigen of interest, will also find use in the subject immunoassays. A number of antibody fragments are known in the art which comprise antigen-binding sites capable of exhibiting immunological binding properties of an intact antibody molecule. For example, functional antibody fragments can be produced by cleaving a constant region, not responsible for antigen binding, from the antibody molecule, using *e.g.*, pepsin, to produce $F(ab')_2$ fragments. These fragments will contain two antigen binding sites, but lack a portion of the constant region from each of the heavy chains. Similarly, if desired, Fab fragments, comprising a single antigen binding site, can be produced, *e.g.*, by digestion of polyclonal or monoclonal antibodies with papain. Functional fragments, including only the variable regions of the heavy and light chains, can also be produced, using standard techniques such as recombinant production or preferential proteolytic cleavage of immunoglobulin molecules. These fragments are known as Fv. See, *e.g.*, Inbar *et al.* (1972) Proc. Nat. Acad. Sci. USA 69:2659-2662; Hochman *et al.* (1976) Biochem 15:2706-2710; and Ehrlich *et al.* (1980) Biochem 19:4091-4096.

A single-chain Fv ("sFv" or "scFv") polypeptide is a covalently linked V_H-V_L heterodimer which is expressed from a gene fusion including V_H- and V_L-encoding genes linked by a peptide-encoding linker. Huston *et al.* (1988) *Proc. Nat. Acad. Sci. USA* 85:5879-5883. A number of methods have been described to discern and develop chemical structures (linkers) for converting the naturally aggregated, but chemically separated, light and heavy polypeptide chains from an antibody V region into an sFv molecule which will fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. See, *e.g.*, US Pat. Nos. 5,091,513, 5,132,405 and 4,946,778. The sFv molecules may be produced using methods described in the art. See, *e.g.*, Huston *et al.* (1988) Proc. Nat. Acad. Sci. USA 85:5879-5883; US Pat. Nos. 5,091,513, 5,132,405 and 4,946,778. Design criteria include determining the appropriate length to span the distance between the C-terminus of one chain and the N-terminus of the other, wherein the linker is generally formed from small hydrophilic amino acid residues that do not tend to coil or form secondary structures. Such methods have been described in the art. See, *e.g.*, US Pat. Nos. 5,091,513, 5,132,405 and 4,946,778. Suitable linkers generally comprise polypeptide chains of alternating sets of glycine and serine residues, and may include glutamic acid and lysine residues inserted to enhance solubility.

"Mini-antibodies" or "minibodies" will also find use with the present invention. Minibodies are sFv polypeptide chains which include oligomerization domains at their C-termini, separated from the sFv by a hinge region. Pack *et al.* (1992) Biochem 31:1579-1584. The oligomerization domain comprises self-associating a-helices, *e.g.*, leucine zippers, that can be further stabilized by additional disulfide bonds. The oligomerization domain is designed to be compatible with vectorial folding across a membrane, a process thought to facilitate in vivo folding of the polypeptide into a functional binding protein. Generally, minibodies are produced using recombinant methods well known in the art. See, *e.g.*, Pack *et al.* (1992) Biochem 31:1579-1584; Cumber *et al.* (1992) J. Immunology 149B: 120-126.

Production of SARS Antigens

The SARSV antigens used in the present invention are generally produced recombinantly. Thus, polynucleotides encoding SARSV antigens for use with the present invention can be made using standard techniques of molecular biology. For example, polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. Furthermore, the desired gene can be isolated directly from viral nucleic acid molecules, using techniques described in the art, such as those described for HCV in Houghton *et al.*, US Pat. No. 5,350,671. The gene encoding the antigen of interest can also be produced synthetically, rather than cloned. The molecules can be designed with appropriate codons for the particular sequence (preferably optimum codons for the expression host of choice). The complete sequence is then assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, *e.g.*, Edge (1981) *Nature* 292:756; Nambair *et al.* (1984) *Science* 223:1299; and Jay *et al.* (1984) *J. Biol. Chem.* 259:6311.

Thus, particular nucleotide sequences can be obtained from vectors harboring the desired sequences or synthesized completely or in part using various oligonucleotide synthesis techniques known in the art, such as site-directed mutagenesis and polymerase chain reaction (PCR) techniques where appropriate. See, *e.g.*, Sambrook, *supra*. In particular, one method of obtaining nucleotide sequences encoding the desired sequences is by annealing complementary sets of overlapping synthetic oligonucleotides produced in a conventional, automated polynucleotide synthesizer, followed by ligation with an appropriate DNA ligase and amplification of the ligated nucleotide sequence via PCR. See, *e.g.*, Jayaraman *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:4084-4088. Additionally, oligonucleotide directed synthesis (Jones *et al.* (1986) *Nature* 54:75-82), oligonucleotide directed mutagenesis of pre-existing nucleotide regions (Riechmann *et al.* (1988) *Nature* 332:323-327 and Verhoeyen *et al.* (1988)

Science 239:1534-1536), and enzymatic filling-in of gapped oligonucleotides using T4 DNA polymerase (Queen *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:10029-10033) can be used under the invention to provide molecules having altered or enhanced antigen-binding capabilities, and/or reduced immunogenicity.

Once coding sequences have been prepared or isolated, such sequences can be cloned into any suitable vector or replicon. Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. Suitable vectors include, but are not limited to, plasmids, phages, transposons, cosmids, chromosomes (including artificial chromosomes, such as BACs or YACs) or viruses which are capable of replication when associated with the proper control elements.

The coding sequence is then placed under the control of suitable control elements, depending on the system to be used for expression. Thus, the coding sequence can be placed under the control of a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator, so that the DNA sequence of interest is transcribed into RNA by a suitable transformant. The coding sequence may or may not contain a signal peptide or leader sequence which can later be removed by the host in post-translational processing. See, *e.g.*, US Pat. Nos. 4,431,739; 4,425,437; 4,338,397.

In addition to control sequences, it may be desirable to add regulatory sequences which allow for regulation of the expression of the sequences relative to the growth of the host cell. Regulatory sequences are known to those of skill in the art, and examples include those which cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Other types of regulatory elements may also be present in the vector. For example, enhancer elements may be used herein to increase expression levels of the constructs. Examples include the SV40 early gene enhancer (Dijkema *et al.* (1985) *EMBO J.* 4:761), the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus (Gorman *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:6777) and elements derived from human CMV (Boshart *et al.* (1985) *Cell* 41:521), such as elements included in the CMV intron A sequence (US Pat. No. 5,688,688). The expression cassette may further include an origin of replication for autonomous replication in a suitable host cell, one or more selectable markers, one or more restriction sites, a potential for high copy number and a strong promoter.

An expression vector is constructed so that the particular coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the control sequences being such that the coding sequence is transcribed under the "control" of the control sequences (*i.e.*, RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence). Modification of the